Development of Animal Cell Engineering Technology Based on Synthetic Biology

鄭, 飛揚

出版情報：Kyushu University, 2022, 博士（システム生命科学）, 課程博士
バージョン：
権利関係：

Development of Animal Cell Engineering Technology Based on Synthetic Biology

Doctoral Thesis

By

Feiyang Zheng

TABLE OF CONTENTS

Abstract

Synthetic biology aims to study genetic elements in cells, synthesize and recombine existing genetic elements to construct intracellular artificial genetic networks to establish genetically engineered organisms. Gene delivery technology is one of the most important tools for establishing engineered cells with artificial gene circuits. Synthetic gene delivery elements can be used to efficiently integrate DNA fragments into the genome of host cells. Various types of gene delivery strategies have emerged for the generation of recombinant cells. However, the gene delivery for gene therapy and generation of recombinant cells tools are still limited. In this study, a novel LINE-1 retrotransposon-based transgene delivery vector was established. LINE-1 is a type of retrotransposon that widely exists in human genome. The LINE-1-based gene delivery tool was proven to induce the transgenes into random sites in the genome. LINE-1-based gene transfer vectors achieved high efficiency of gene transfer in CHO-K1 cells by the “copy-and-paste” principle, compared with the other gene delivery methods. The retrotransposon vector contains a neomycin resistance gene split by an intron as a marker gene, and a gene encoding an antibody single-chain variable fragment (Fv) fused with the constant antibody region (Fc) (scFv-Fc) as a model target gene. G418-resistant Chinese hamster ovary cells were generated using this retrotransposon vector, and scFv-Fc was produced in the culture medium. To regulate retrotransposition, we developed a retrotransposon vector system that separately expressed the two open reading frames (ORF1 and ORF2) of LINE-1. Genomic PCR analysis detected the transgene sequence in almost all tested clones. Compared with clones established using the intact LINE-1 vector, clones generated with the split ORF1 and ORF2 system showed similar specific scFv-Fc productivity and retrotransposition efficiency. This study provides a new tool for the establishment of genetically modified cells.

RNA expression contains a wide range of information including physical conditions, chemical environments and endogenous signal. Regulating gene expression

can manipulate cell fate, induce biopharmaceutical production, or modulate tissue morphology. Therefore, gene expression regulation by detecting RNA level can regulate cell fate according to the state of the cells. An artificial transcriptional activation system for RNA detection was developed in this study. To achieve RNA-induced transcriptional activation, we constructed expression plasmid vectors encoding MS2 bacteriophage coat protein fused with transcriptional activator and PP7 bacteriophage coat protein fused with tetracycline repressor. A target RNA encoding MS2 and pp7 aptamers and poly(U) terminal sequence was expressed under the control of human U6 promoter. Recombinant CHO cells harboring a GFP expression unit driven by a TRE promoter were established. The reporter cells were transfected with the vectors encoding fusion proteins and target RNA containing MS2 and pp7 sequences. GFP expression was observed in the cells transfected with all required vectors. In the condition without target RNA, GFP expression was not observed at all. GFP mRNA was upregulated over 30-fold compared with that without target RNA expression vector. Construction of RNA detection systems was also possible using other combinations of RNA motifs and proteins such as com/COM and BoxB/N22. To detect different RNA molecules simultaneously, a target RNA expression vector encoding BoxB and pp7 sequences and a Gal4-N22 fusion protein expression vector were constructed. Recombinant CHO cells harboring TRE-GFP and UAS-RFP expression cassettes were co-transfected with the fusion protein expression vector and MS2-pp7- or MS2-BoxB-tagged RNA expression vector. The expression of GFP or RFP was induced in the cells when corresponding target RNA vector was transfected without cross talk. In addition, cells expressing both MS2-pp7- and MS2-BoxB-tagged target RNAs exhibited the inducible expression of both GFP and RFP, as expected. The RNA detection systems have the potential to provide new gene regulatory systems that direct gene expression cascades.

Biopharmaceuticals such as recombinant proteins produced by Chinese hamster ovary (CHO) cells are extensively used. Reduced culture temperature provides an efficient benefit to improve the productivity and viability by decreasing the nutrient consumption, indicating the shift of energy from cell growth toward product synthesis.

A cold-inducible gene expression system was developed using artificial genetic circuit based on trans-activator system. To regulate the gene expression depends on temperature shift, a cold-shock promoter mediated cold-inducible trans-activator expression system, a tetracycline expression regulation system-based positive feed-back amplification system and a scFv-Fc expression system have been developed for inducing temperature shift responsive transgene expression, enhancing transgene expression and initiating biopharmaceutical production, respectively. The combining of these three systems achieved cold-inducible high-level production of scFv-Fc. This engineered gene expression regulation system has the potential to increase targeted drug production. Higher level expression of reporter gene was confirmed in lower temperature using cold-inducible trans-activator expression system. A temperature shift responsive scFv-Fc expression was confirmed when cold-inducible trans-activator expression system was combined with scFv-Fc expression system. High level expression of scFv-Fc was observed in cells with both cold-inducible trans-activator expression system, positive feed-back amplification system and scFv-Fc expression system. This study showed that CHO cells with the cold-inducible transactivator expression can boost the target protein productivity.

In summary, a serious of synthetic gene elements were developed for establishing genetic engineered organisms. Both of them have the potential to be applied for tissue engineering, gene therapy or biopharmaceutical production.

Chapter 1
Background

1.1 Synthetic biology

Synthetic biology seeks to synthesize artificial gene element in living cells for generating genetically modified organisms with natural and synthetic biological elements, which is a developing technique for tissue engineering, gene therapy and biopharmaceutical production **[1-4]**. Synthetic gene devices with a wide range of functions have been developed, including accumulative gene integration system for scFv-Fc antibody production, heat-inducible transcriptional regulation system for liver function regeneration **[5]**, cell type responsive expression system for gene therapy **[6]** and logic gates signal processing system for controlling metabolic activities **[7-9]** (Figure 1).

Mammalian cells are complex machines with highly sophisticated regulatory networks. The natural genetic elements found within these cells work in concert to precisely control cellular functions, including cell cycle, DNA replication, chromosome recombination, gene expression, energy generation, extracellular matrix secretion, tissue structure, organ function, and the overall appearance of organisms. Artificial gene elements, based on natural genetic elements, have been created to artificially regulate various aspects of life, from cell fate to physical form **[10, 11]**.

Since Francis Crick and James Watson discovered the DNA double helix structure and proposed the central dogma of molecular biology **[12]**, a new understanding of the regulatory mechanism was generated. Central dogma of molecular biology describes the transfer of information stored in biopolymers in living organisms **[13]**. Based on the information transfer, genetic element derived biopolymers, including DNA, RNA and proteins, collectively regulate cellular functions.

Synthetic gene elements are engineered through the assembly and modification of natural elements. Due to the shared central dogma of all organisms, genetic elements from one organism can function in another. As a result, genetic elements from various

organisms have been studied and incorporated into other organisms to manipulate cell functions and control life activities. This approach led John Craig Venter and his team to create a bacterium with a completely chemically synthesized genome, which was capable of replicating and forming colonies **[14]**. Synthetic biology typically constructs DNA, RNA or protein devices for changing the regulatory mechanism in cells to achieve specific functions **[15]**.

In this study, a recombinant cell line establishing strategy, aiming to build a variety of gene modules for gene delivery or gene expression regulation was developed.

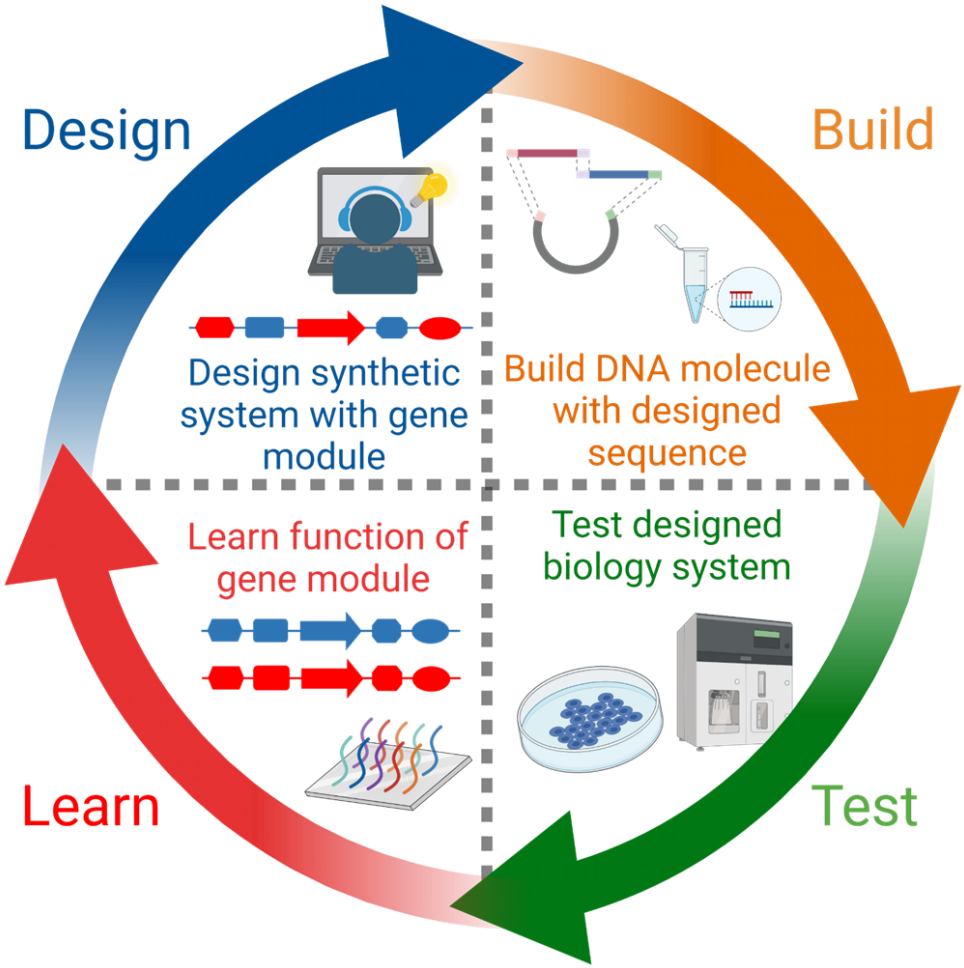


Figure 1. Synthetic biology seeks to synthesize artificial element in living cells for generating recombinant cells with natural and synthetic biological elements **[16]**.

1.2 Gene delivery technology

Gene delivery procedures is one of the important tools for the generation of genetically modified organisms, which have been applied for tissue engineering, gene therapy or biopharmaceutical protein production **[17]**. Compared with random integration, the synthetic gene delivery vector using synthetic biology has higher delivery efficiency and a variety of functions. Various types of gene delivery methods have been developed to efficiently integrate transgenes (Table 1).

Transposon elements-based gene delivery strategy is one of the most widely used gene delivery method. Transposition mediated by transposon can transfer DNA fragment to a random location in the genome by a process called "cutting and pasting" without RNA intermediate. PiggyBac is a type of transposon derived from Lepidoptera insects which have been applied for developing an efficient non-viral gene delivery method. It was first discovered and isolated when baculovirus invaded the *Spodoptera litura* cell line TN-368 **[18]**. PiggyBac is an autonomous factor with a length of 2,476 bp, a short terminal inverted repeat (ITR) and an open reading frame (ORF). The ORF fragment in Piggybac is coding a nuclease named PBase which can cut DNA sequence between ITRs and integrate it into random site in genome. Transposon based gene delivery vector is suitable integration tool for high expression of the target gene. The engineered PiggyBac binary vector system can carry DNA sequences up to 15 kb for transposition **[19]**.

The development of gene delivery technology using viral vectors is widespread. Unlike non-viral gene delivery methods, viral vectors have the ability to infect cells autonomously, allowing the delivered nucleic acid to enter cells without the need for transfection methods such as electroporation. Several types of viral vectors, including adenovirus (AdV), adeno-associated virus (AAV), and lentivirus, have been utilized for the creation of genetically modified organisms **[20-22]**. The Lentiviral vector is a highly efficient tool for gene delivery, commonly used for generating recombinant cells and implementing gene therapy. It is developed using the HIV-1 virus as a basis. Lntiviral vectors are designed to retain non-coding cis-elements such as long-terminal repeats

(LTR) and include trans-elements such as the Gag-polypeptide required for packaging. The HIV-1 envelope protein is replaced by either murine leukemia virus (MLV) amphotropic envelope protein or VSV-G to increase the range of cells that can be infected. To prevent the reformation of a replicable virus, the transgene elements of the lentiviral vector are dispersed across three or more plasmids. **[23-25]**. Target gene is transcribed into RNA and packaged into virus. The packaged virus efficiently infects target cells and integrates the target gene into the genome through reverse transcription.

On the other hand, a targeted integration method using Cre recombinase has also been developed. Discovered in 1981 from the P1 phage, Cre recombinase belongs to the lambda Int enzyme superfamily. It consists of a 1,029 bp coding region that encodes a 38 kDa monomer protein with 343 amino acids. Cre can trigger recombination between two loxP sites. **[26]**. Therefore, gene fragment can be integrated from donor vector to target loxP site in genome. An accumulative site-specific gene integration system (AGIS) using Cre-recombinase and mutated loxP sites have been developed for establishing antibody high production CHO cell lines **[27]**.

Recently, with the development of genome editing technology, programmable nuclease such as CRISPR, TALEN, Zinc finger nuclease are being widely used to establish recombinant cell lines. Target cleavage site of nuclease could be designed by changing the arrangement of amino acid sequence in DNA binding protein or base sequence in guide RNA. Thus, the targeted integration of transgene can be accomplished through homologous recombination repair after inducing double-strand breaks with the use of programmable nucleases, without the dependence on specific sequences such as loxP. **[28]**. Although delivering multiple copies of transgene is still challenging for genome editing tools, CRISPR system is widely used due to its programmability **[29]** (Figure 2).

However, it is still challenging to develop a gene delivery method with high efficiency, high load and low immunogenicity.

Table 1 Gene delivery system developed for recombinant cell generation.

Delivery vector	integration mechanism	integration attribute	Reference
Lentivirus	Reverse transcription	Random	**[23-25]**
PiggyBac transposon	DNA recombination	Random	**[19]**
Cre/loxP	DNA recombination	Site specific	**[27]**
CRISPR/Cas9	Homology directed repair	Site specific	**[28]**

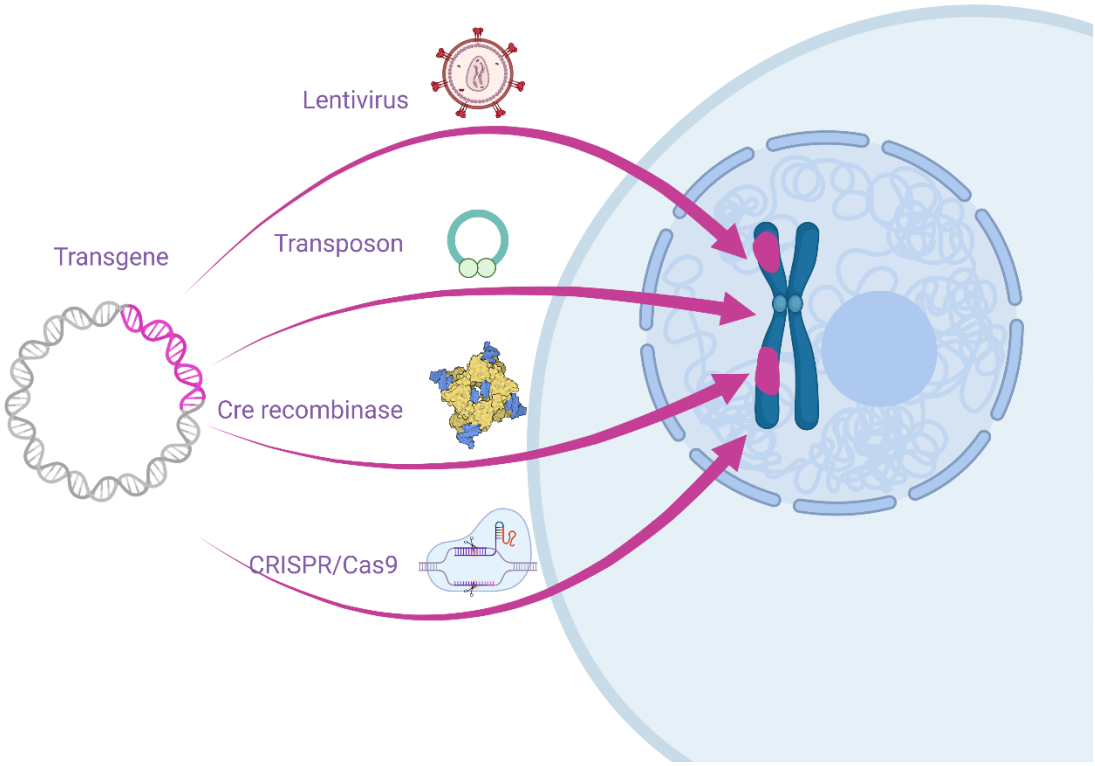


Figure 2. Schematic representation of typical gene delivery method.

1.3 Gene expression regulation

Synthetic biology components, including those that regulate gene expression, have been developed to control cell functions and process signals. Both natural and synthetic elements can be used to engineer artificial biological systems that control gene expression by regulating gene transcription, RNA translation, or protein function. These synthetic gene networks mimic the functions of natural gene networks by processing signals and controlling cell fate through regulation of gene expression. These gene regulation systems, which process signals like an electronic circuit, are referred to as gene circuits.

To engineering artificial expression regulation system, nature gene components was discovered and optimized for assembling synthetic gene elements. Various of gene circuits have been developed for controlling cell function. IPTG inducible gene expression regulation system was developed and widely used for recombinant protein production and colonies selection in bacteria. Although protein expression in bacteria and animal cells is based on almost the same central dogma, their regulatory systems are hardly universal due to the differences of expression systems. Therefore, engineered expression regulation elements are still not enough for animal cells when compared with the huge demand for developing controllable recombinant cells.

Transcriptional regulation plays a critical role in controlling gene expression in animal cells, thereby influencing various cellular processes, especially cell differentiation. The study of gene regulation at the DNA level has led to the establishment of a new field called epigenetics. Numerous transcriptional regulation elements have been discovered in the animal genome and are widely used in synthetic biology. Various gene expression systems, such as the tetracycline and lactose expression regulation systems, have been developed **[30]**. The study of transcriptome also provides a great potential to discover new transcriptional regulatory elements with specific functions.

The development of artificial transcriptional regulation systems involves the discovery, modification and assembly of natural elements. Two types of elements are

found in nature cells affecting the gene transcription, cis-acting elements and trans-acting elements. Cis-acting elements including promoters, enhancers and operators, are a type of DNA sequence that affect the transcription of genes nearby these elements without coding protein. Cis-acting elements are recognized by transcriptional regulating proteins or RNA polymerase for controlling gene expression. The distance between cis-acting elements and regulated gene influences the effect of cis-acting elements. Trans-acting elements, on the other hand, are a type of open reading frame that code the transcriptional regulating proteins. Transcriptional regulating proteins could targeted recognize cis-acting elements and regulate the expression of genes nearby its target. The distance between trans-acting elements and regulated gene has not influence on the effect of trans-acting elements. Based on these elements, a variety of tight and efficient gene expression regulation systems have been developed. In recent years, genome editing technology has further expanded the applicable range of transcriptional regulation. Thus, transcriptional regulation is a type of typical biological signal for gene circuit construction.

By engineering gene expression regulation system, gene circuits can be designed to achieve specific functions, sensing specific environmental stimuli, processing signals and controlling cell's fate **[31, 32]**.

Engineered controllable gene expression systems were created by incorporating responsive elements based on natural sensitive biomolecules that can be activated by specific stimuli, thus regulating cell function through gene expression regulation. Physical stimulations responsive elements have been developed for constructing genetic switches that can artificially control gene expression by altering physical condition. Temperature sensors have been applied as a production regulator in industrial fermentation and a trigger for gene therapy **[33, 34]**. Light with different wavelength is also used as a fast and precise gene expression trigger by engineering light sensor. The light sensor developing technology have been termed "Optogenetics" **[35, 36]**. Magnetic field **[37]** sensor and ultrasound wave **[38]** sensors have been developed as a oncolysis methodology. Sensors for chemical stimulations were usually developed for constructing recombinant cells that can automatically respond

to changing environments to adapt to different environments. Tissue-engineered artificial tissues have been designed to accelerate vascular network formation under hypoxia using oxygen sensor **[39]**. Ion concentrations can reflect the electrical activity state of nerve or muscle cells, ion sensors have been developed for neurobiological research **[40, 41]** Antibiotic resistance gene expression regulation systems based on antibiotic sensors was evolved in bacteria due to the widely using of antibiotics. Engineered antibiotic sensors including tetracycline expression control systems have become one of the most widely used gene expression control systems **[42, 43]**. Endogenous signal sensors including DNA damage sensor **[11]** and ligand signal sensor **[31]** have also be emerged for controlling cells fate. However, the developing of practical synthetic network remains challenging due to the lack of efficient and composable genetic components. Therefore, novel gene elements with high sensitivity, high dynamic range, and high composability can facilitate the development of gene circuits.

To develop signal processing system, a type of biological system that engineered for processing specific signal, can alter signals generated by sensors for achieving specific functions. The most common used method for signals processing is transcriptional regulation. To implement basic logical judgments, intracellular logic circuits based on transcriptional activation have been constructed. Logic circuits including OR, NAND, NOR and XOR logic gate using drug inducible transcriptional regulation system and recombined promoter have been introduced into mammalian cells **[44]**. The affinity of DNA binding domains to specific DNA element were controlled by drugs and transcriptional activation domains were fused with DNA binding domains for regulating transcription. In bacteria cells, on the other hand, gene circuit including AND, OR, NOR, and NAND gates using aptamer ligand inducible ribozyme cleavage system have been developed **[32, 43, 45]**. The development of this synthetic biology system demonstrates the feasibility of developing complex algorithms inside cells. To switch cells states, toggle switch gene circuit have been developed. Lebar T. et, al. constructed a switchable transcriptional regulation system based on transcription-activator-like effectors (TALE)-repressor-based genetic circuit.

Two protein expression units that suppress each other controlled by drug inducible expression regulation system results in switchable epigenetic bistability, which achieved an intracellular memory **[46, 47]**. These synthetic systems have the potential to regulate cell fate or production of biopharmaceuticals. To develop engineered organisms inside cells for timekeeping functions, oscillating network in bacteria have been developed. Three sequential inducible expression unit were designed to repress each other and periodically induces the synthesis of green fluorescent protein (Table. 2).

The design of regulatory genes is crucial for achieving specific functions. Gene circuits have been developed to induce revascularization in order to increase the viability of transplanted tissues in vivo . This gene circuit controls the expression of vascular endothelial-derived growth factor (VEGF), leading to the formation of a capillary network in tissues with low blood flow. **[39]**. To develop oncolytic viruses that attack specific cancer cells, an engineered virus that can replicate selectively was developed **[4]**. Genes that necessary for adenovirus replication was controlled by the synthetic gene circuit for directed replication in cancer cells.

Table 2 Gene expression regulation systems.

Gene module	application	Reference
Light sensor	Cell fate controlling	**[35, 36]**
Oxygen sensor	Tissue engineering	**[39]**
Heat sensor	Gene therapy	**[37]**
Ion sensor	Neurobiological study	**[40, 41]**
Logic gate	Gene therapy	**[44]**
Toggle switch	Cell reprogramming	**[46, 47]**
Virus assemble switch	Gene therapy	**[4]**
Blood vessel regeneration	Tissue engineering	**[39]**

1.4 Outline of this thesis

Various types of synthetic biology elements have been developed to achieve different functions in mammalian cells. However, the synthetic biology components applicable to mammals are still limited. The aim of this study is to create a series of new synthetic biology components that expand the scope of synthetic biology in mammalian cell engineering.

Delivery of multiple copies of genes can effectively enhance transgene expression levels and improve the functionality of engineered cells. The challenge in conventional gene delivery techniques is that they do not efficiently amplify the copy number the introduced gene, limiting the effectiveness in boosting the expression levels and functional output of engineered cells. LINE-1 retrotransposon has been discovered from genome of animal cells as a independent moveable element. In order to develop a gene delivery solution that can amplify the copy number of gene, the LINE-1 transposon was engineered by assembling the LINE-1 transposon with the artificial gene expression element. In this study, a LINE-1 transposon based synthetic transgene delivery vector was firstly developed.

Numerous gene circuits with various functions have been created, but few gene expression regulation systems have been developed specifically for animal cells. To expand the application of gene circuits, a novel artificial transcriptional activation system for RNA detection was developed in this study.

Cold or IPTG inducible recombinant protein production system have been developed and widely applied in *E. coil* cells. Recombinant proteins produced by mammalian cells have been applied as biopharmaceuticals. However, few inducible expression systems were applied for protein production in mammalian cells. Therefore, in this study, a cold inducible gene expression system was designed based on trans-activator for biopharmaceutical production.

Chapter 2

Development of a retrotransposon-based gene delivery system

2.1 Introduction

Gene delivery is an essential step in creating genetically modified cells by introducing DNA fragments into the cells. Over time, various types of gene delivery methods were emerged to facilitate the efficient integration of transgenes **[17, 48]**. Gene transfer strategy integrating multi-copy of gene with high efficiency may benefit the target gene expression. However, gene delivery technology that can amplify copy number of transgene is still limited.

To tackle the limitations of conventional gene delivery methods, a novel vector was designed that can not only introduce the desired gene into the genome but also amplify its copy numbers through reverse transcription utilizing the LINE-1 retrotransposon technology.

Mobile elements occupy about 45% of the human genome. Two types of mobile elements, retrotransposons and transposons, were discovered according to the manner of transposition. Retrotransposons create a copy of DNA by reverse transcription and transposition within the genome by the principle of "copy and paste", in which the copy of retrotransposon is inserted into genome and the number of copies is amplified each time transposition occurs. Transposons, on the other hand, can transposition within the genome by the “cut and paste” principle, in which transposons themselves are excised from the genomic DNA and integrated into random DNA site. The copy number of transposon does not change before and after transposition **[18]**.

Retrotransposons are a category of mobile genetic elements that move within a genome through the process of transcription and reverse-transcription. They are classified into two types: long-terminal repeat (LTR) and non-LTR retrotransposons. LTRs type retrotransposons are similar to retrovirus and have long terminal repeat

sequences at both ends of retrotransposon. Non-LTR type retrotransposons, on the other hand, have no LTRs. Non-LTR retrotransposons are the most abundant retrotransposons in mammals and can be further classified into long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE) **[49, 50]**. LINE consists of a sequence of 5,000-7,000 bp in total length and encodes proteins essential for transposition, whereas SINE consists of a 90-300 bp sequence and does not encode any protein essential for transposition. Therefore, LINE-derived proteins are required for the transposition of SINE. The SINE promoter has an internal promoter for recruiting RNA polymerase III, whereas the LINE retrotransposon is primarily transcribed by RNA polymerase II.

LINE-1 is the most abundant LINE in human cells **[51]**. LINE-1 accounts for 17% of DNA in humans. However, most LINE-1s are retrotransposition defected that cannot induce retrotransposition due to internal rearrangements or 5'-terminal truncations. Only 80-100 copies of LINE-1 are capable of retrotransposition among millions copies of human LINE-1. Full-length LINE-1 has approximately 6 kbp and contains a 5'UTR with promoter activity in both the sense and antisense orientations, two ORFs and a 3′ UTR terminating with an AATAAA polyadenylation signal. ORF1 encodes the p40 protein with RNA-binding and chaperone activity, while ORF2 encodes a 150 kDa protein that have endonuclease and reverse transcriptase activity. During LINE-1 retrotransposition, ORF1 and ORF2 proteins bind to their own RNA in the cytosol and form ribonucleoprotein particles (RNPs), which facilitate retrotransposition of LINE-1 RNA. Various of research for discovery the mechanism of LINE-1 have been reported **[52]**. However, few study is emerged for developing LINE-1 as a gene delivery tool in mammalian cells (Figure 3).

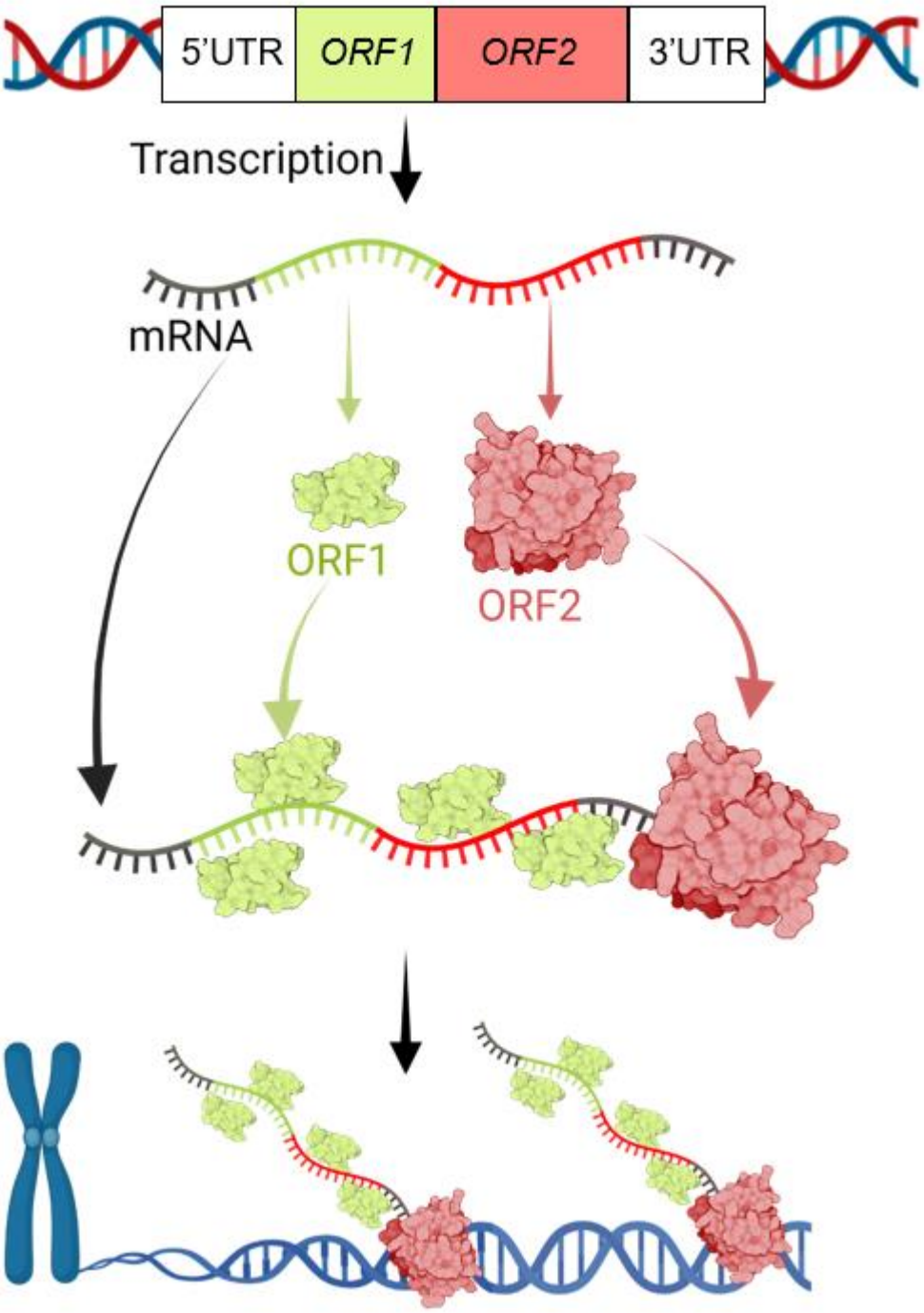


Figure 3. Schematic representation of LINE-1 retrotransposon.

Varity of gene delivery method based on DNA transposon have been developed. Recombinase generated by the coding sequence in DNA transposon could initiate the DNA recombination on the untranslated region of DNA transposon. Based on DNA transposon, gene delivery tools have been developed, in which gene of interest was

transferred with the help of recombinase. High integration efficiency was observed and the delivery method was widely applied for recombinant cells or genome modified animal generation. To regulate the transposition, the nuclease of transposon that is essential for transposition was provided separately. Dual plasmid gene delivery strategy was emerged for preventing unexpected transposition.

Few gene transfer tools based on retrotransposon were reported, although serious of gene delivery tools based on transposons have been emerged. Retrotransposons are transposons that replicate themselves by reverse transcription using their own mRNA as a template. LINE-1 is a common retrotransposon in the human genome. ORF2 and ORF1 proteins encoded by LINE-1 coding sequence can induce the LINE-1 self-replication and transfer by reverse transcription. In the final stage of retrotransposition, a copy of the original LINE-1 retrotransposon is inserted at a new site in genomic DNA. Therefore, retrotransposons are attractive gene transfer tools that can be applied for delivering transgene and increasing the copy number of a target gene **[29]**. In this study, retrotransposon-based gene elements were engineered as a gene delivery tool for genomic organs generation.

Biopharmaceuticals are molecules produced by living organisms. Recombinant proteins, one of widely used biopharmaceuticals, are produced by introducing target foreign gene into host cells. However, recombinant proteins require complex structures and appropriate post-translational modifications. Therefore, animal cells such as CHO cells, which are capable of exact post-translational modification and high level protein expression, are generally used for biopharmaceuticals production. Therefore, in order to enhance the production level in animal cells, attempts have been made to construct highly productive engineered cells by amplifying the target gene on the chromosome.

Antibodies, also known as immunoglobulins, is parts of immune system that remove pathogens invaded the body and destroy infected cells or cancer cells. When antigen enters the body, through a series of immune responses, B lymphocytes differentiate into plasma cells and secrete antibodies that can specifically bind to the antigen for capturing and removing antigen **[53]**. Artificial engineered antibody is a type of biopharmaceuticals that widely used as targeted drugs.

Single-chain variable fragment (scFv) is an immunoglobulin that fused the variable regions of the heavy and light chains with a short linker peptide. The protein still retains the essence of the original immunoglobulin. A human IgG Fc fragment was fused to an anti-prion scFv to generate scFv-Fc. A recombinant CHO-K1 cell line with the productivity of 60 pg/cell•day using accumulative site-specific gene integration system (AGIS) have been established **[54, 55]**.

In this study, a LINE-1 vectors containing LINE-1 component, drug resistance gene and antibody producing gene inserted in 3'UTR for delivery *scFv-Fc* gene was developed (Figure 4). A typical antibody gene, *scFv-Fc*, was introduced into CHO-K1 cells using LINE-1 based gene delivery method.

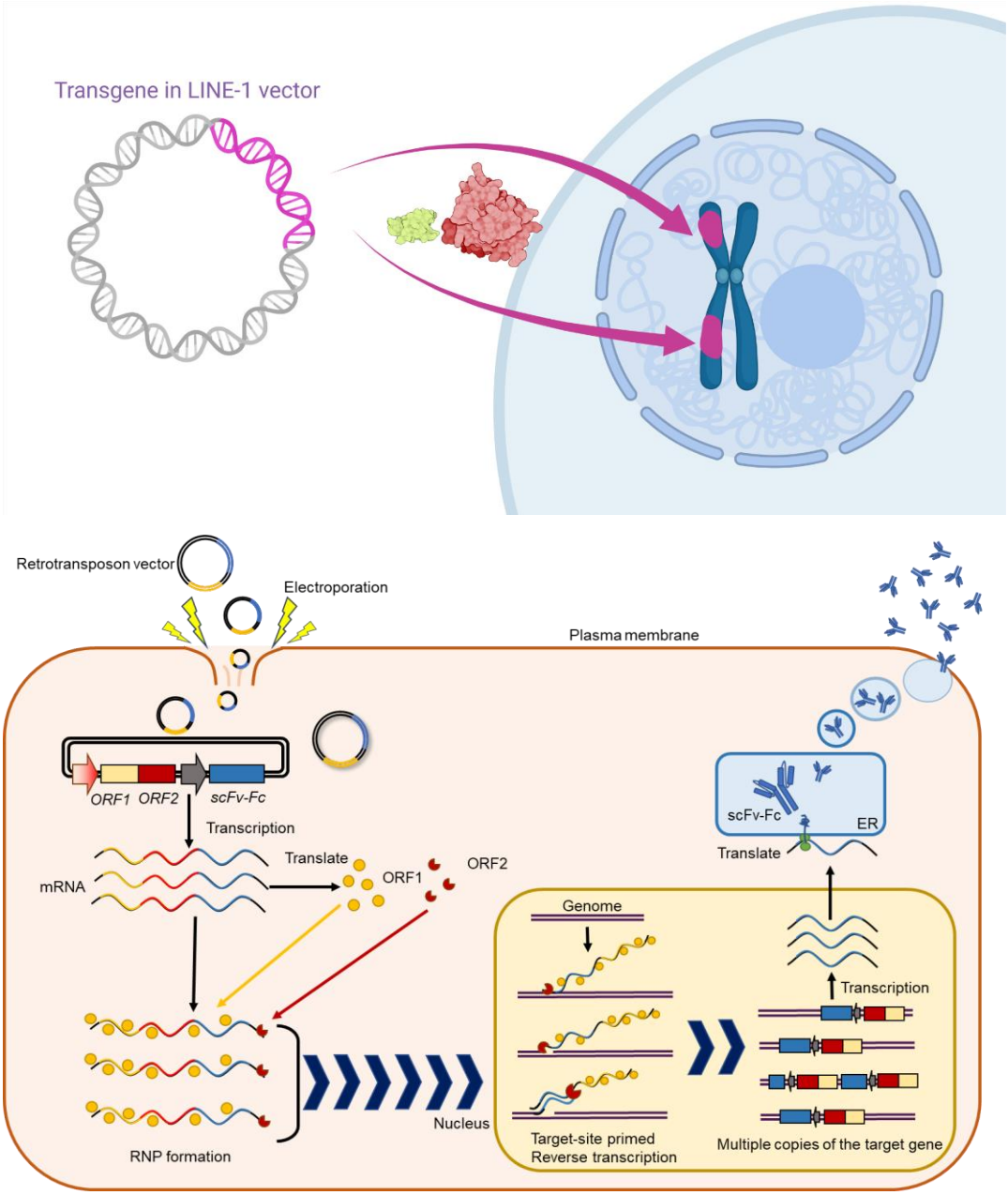


Figure 4. Schematic representation of transgene delivery mediated by LINE-1 vector

2.2 Materials and methods

Cell culture

CHO-K1 cells (RIKEN, Tsukuba, Japan) were cultured in CHO-K1 culture medium containing F-12 Ham (Sigma-Aldrich, St. Louis, MO, USA) with 10% fetal bovine serum (BioWest, Nuaillé, France) with non-treated cell culture plates (Thermo Fisher Scientific) at 37°C and 5% (v/v) CO_2. To prevent the contamination, the culture medium was added with 90 μg mL^{-1} penicillin G potassium (Fujifilm Wako Pure Chemical Industries, Osaka, Japan) and 100 units mL^{-1} streptomycin sulfate (Fujifilm Wako).

Plasmid construction

Mouse derived LINE-1 vector were prepared by chemically synthesized (Genewiz, South Plainfield, NJ, USA). A TK promoter controlled human γ-globin intron interrupted neomycin resistance reporter gene expression unit a was inserted at reverse direction into the 3′-UTR region of LINE-1 vector. A CMV promoter controlled *scFv-Fc* expression unit was inserted into downstream of the *Neo*^{r} gene to generate a transgene delivery vector based on LINE-1. The recombinant vector, pLINE1-scFvFc, was prepared as intact type LINE-1 vector plasmid for evaluating the transgene vector (Figure 5).

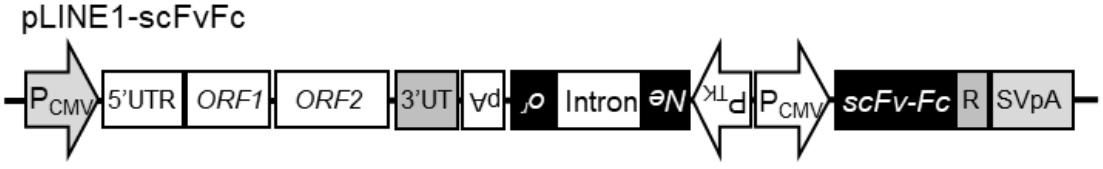


Figure 5. Structure of LINE-1 based transgene delivery vector.

To regulate gene transfer of LINE-1 based gene transfer system, the coding sequence were removed from pLINE1-scFvFc. The ORF1 in LINE-1 delivery vector was replaced by point-mutated ORF1 gene (Genewiz) to generate ORF1 mutated LINE-1 vector. To construct the ORF1 and/or ORF2 expression vectors, High accuracy PCR was performed for extracting ORF1, and ORF2 gene from the LINE-1 derived plasmid using KOD plus Neo DNA polymerase (Toyobo, Osaka, Japan). IRES-DsRed DNA fragment derived from pIRES2-DsRedExpress plasmid vector (Clontech, Mountain View, CA, USA) was ligated into pZeoSV2 (Invitrogen) plasmid to generate negative control vector pZeo/IRES-DsRed. To generate pORF1–2, pORF1, and pORF2, the amplicon of ORF1, and ORF2 fragment were ligated into pZeo/IRES-DsRed (Figure 6).

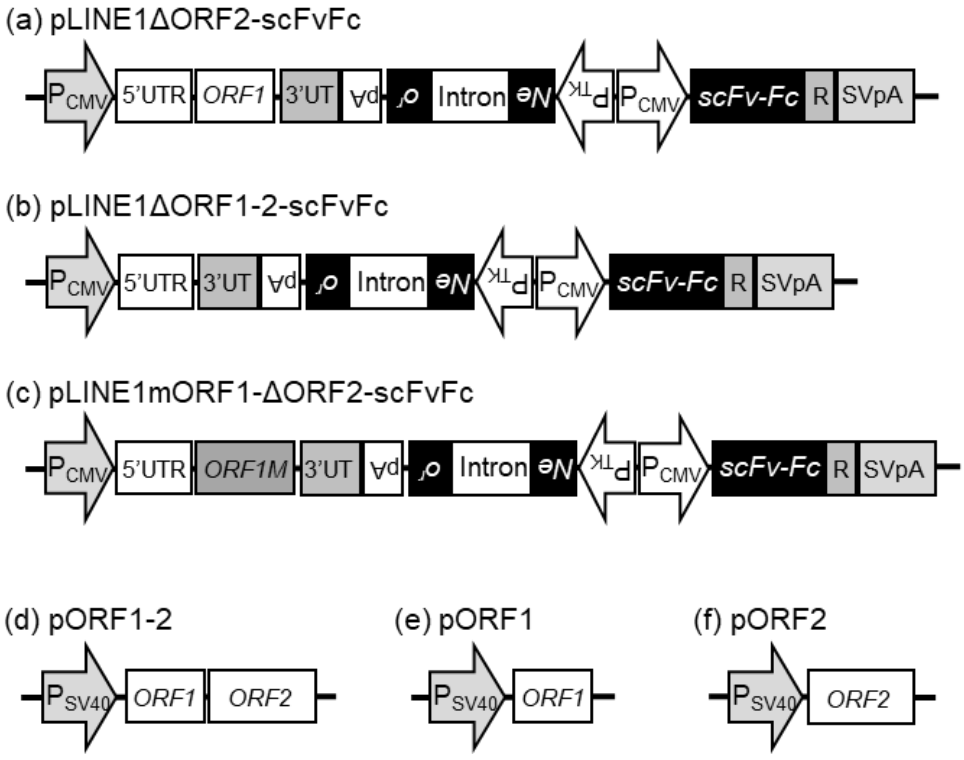


Figure 6. Structure of LINE-1 based ORFs separated providing transgene delivery vector. (a) ORF2 removed LINE-1 vector (b) ORF1 removed LINE-1 vector (c) ORF1 mutated ORF1 removed vector (d) ORF1-2 expression vector (e) ORF1 expression vector (f) ORF2 expression vector.

Retrotransposition assay

Recombinant CHO cells generated by LINE1-based gene delivery tools were established as follows. CHO-K1 cells (10^4 cells) were transfected with plasmids of total weight of 1000 ng. ORF-removed LINE-1 derived vectors and ORF expression vectors were transfected at the ratio of 10:1. The electroporation was performed using the 10μL kit of Neon Transfection System (Invitrogen) at the conditions of three pulses at 1650 V for 10 ms. Transfected cells were seeded in a 24-well plate (Thermo Fisher Scientific).

For caffeine treated cells, the culture medium was replaced with 0.04–4 mM caffeine containing culture medium (Sigma-Aldrich) at 24 h after transfection and cultured for 24 h. Transfected cells were seeded at a seeding density of 2.0×10^4 cells/well in 6-well plates (Thermo Fisher Scientific) 5 days after transfection with 2 mL of 0.8mg G418 (Sigma-Aldrich) containing F-12 culture medium for selection. Cell colonies were counted after phosphate-buffered saline (PBS) washing, 4% paraformaldehyde (Narabyori Laboratory, Nara, Japan) fixing and 0.05% crystal violet (Sigma-Aldrich) staining for retrotransposition efficiency evaluation. The clones were established by picking up cell colonies using PBS and colonies treated with 10 μL 0.25% trypsin-EDTA solution (Thermo Fisher Scientific) 7 days after G418 screening.

Genomic PCR and copy number assessment

Genomic DNA of recombinant cells was extracted by MagExtractor genomic DNA extraction kit (Toyobo). Integrated *Neo*^{r} and *scFv-Fc* genes were confirmed by PCR using G-Taq polymerase (CosmoGenetech, Seoul, South Korea) to determine the function of LINE-1 based gene delivery method.

Real-time PCR was performed using PikoReal Real-Time PCR System (Thermo Fisher Scientific) and quantitative PCR reagent (Thunderbird Probe qPCR Mix; Toyobo) for evaluating the copy number of the *scFv-Fc* gene in the recombinant CHO cell genome. Real-time PCR data was analyzed by PikoReal Software 2.0 (Thermo Fisher Scientific). Recombinant scFv-Fc producer CHO cells (CHO/scFv-Fc x1) was

applied as single-copy control.

ScFv-Fc productivity

The specific scFv-Fc productivity of recombinant cells engineered by LINE-1 derived gene transfer tools was evaluated by enzyme-linked immunosorbent assay (ELISA). Cells were seeded in a 24-well plate (Thermo Fisher Scientific) at a density of 2.5×10^4 cells/well. The culture medium was collected and replaced daily during 4 days culturing. Cells were harvested every day for counting the number of viable cells using the trypan blue. ELISA was performed for measuring the concentration of scFv-Fc in culture medium. Rabbit anti-human IgG (Fc) antibody (#609-4103; Rockland Immunochemicals, Philadelphia, PA, USA) and rabbit peroxidase-conjugated anti-human IgG (Fc) antibody (#609-4303; Rockland Immunochemicals) were used as the primary and secondary antibodies, respectively. A dilution series of the human Fc fragment (Jackson Immuno Research, West Grove, PA, USA) was applied as a standard.

Table 3. PCR method used for developing LINE-1 derived gene transfer tools

PCR	Application	PCR setting
High accuracy PCR	Plasmid construction	98°C for 2 min, 35 cycles of amplification at 94°C for 15 s, 64°C for 30 s, 68°C for 15–45 s.
Genome PCR	Gene detection	95°C for 2 min, 35 cycles of amplification at 95°C for 30 s, 58°C for 40 s, 72°C for 15–70 s and 72°C for 5 min.
Real-time PCR	Gene copy number detection	95°C for 1 min, 40 cycles of amplification at 95°C for 15 s, 60°C for 60 s.

Table 4. Primers used for ORF1 and/or ORF2 expression vectors.

Gene		Sequence (5′→3′)
ORF1-ORF2	FW	AAA AGA TCT ACC ATG GCC AAG GGC AAG CGC CG
	RV	AGC TTC TGC TTC CAA GAT CTG TCC A
ORF1	FW	AAA GCT AGC ACC ATG GCC AAG GGC AAG CG
	RV	AAA GGA TCC TTA GCG GCG GGT CTT CTC CAG G
ORF2	FW	AAA CTC GAG ACC ATG CCC CCC CTG ACC ACC AAG ATC AC
	RV	AAA GGA TCC CTA GTA GCC GCT GAT CAG GCT GTA CAT GTT

Table 5. Primers used for genomic PCR analysis.

Primer	Sequence (5′→3′)
α	TCG CCC AAT AGC AGC CAG TC
β	GAC TGG GCA CAA CAG ACA ATC G
γ	CGA GGC TGT CTA TTT CTG TGG G
δ	ATG AGG GTG TCC TTG GGT TTT
ε	AGCATCGAGAACATCGGCAC
ζ	GGGCTTGCCCTTGTAGGTCA
η	ATC AAG AAC AGC GGC GAC TCG
θ	GAT GAA CAG AGC GGC GAT GA

Table 6 Primers and probes used for real-time PCR analysis.

Gene	Sequence (5′→3′)	
scFv-Fc	FW	TGG AGT GGG AGA GCA ATG G
	RV	CGG GAG GCG TGG TCT TG
	Probe	AGC CGG AGA ACA ACT
GAPDH	FW	TCT GGC AAA GTG GAA GTT GTT G
	RV	CTT GCC ATG GGT AGA GTC ATA CTG
	Probe	CCT CAA CTA CAT GGT CTA CA

Statistical analysis

Statistical comparisons were evaluated using Student's t-test and comparisons with $p < 0.05$ were considered significantly different.

2.3 Results

Transgene retrotransposition mediated by LINE-1 vectors

Retrotransposons are the most abundant mobile elements detected in mammalian genomes. Among them, LINE-1 is one of widespread retrotransposons in the human genome. The LINE-1 retrotransposon has been investigated as a potential gene delivery vector due to its ability to mediate transposition in human cells. The development of LINE-1-baseed vectors has led to the emergence of a method for delivering shRNAi. However, LINE-1 based vector for transgene delivery was still not reported.

In this study, a LINE-1 vectors encoding the LINE-1 component genes and untranslated regions (5′-UTR, ORF1, ORF2, and 3′-UTR) required for the retrotransposition of transgenes was constructed.

To establish a gene delivery method based on retrotransposon, a LINE-1 vector with neomycin resistance gene and scFv-Fc expression unit were chemical synthesized and transfected into CHO-K1 cells by electroporation. The neomycin resistance gene was split by an intron and placed in antisense direction in 3'UTR of LINE-1. Therefore, split neomycin resistance gene could be expressed only when LINE-1 was reverse transcribed with a retrotransposition assay to confirm retrotransposition event. After 7 days of neomycin selection, drug resistance clones were established. Genomic DNA were extracted from clones and integrated DNA sequence was confirmed by genomic PCR. The primer, α and β, was designed for neomycin resistance gene can distinguish between retrotransposition and random integration. The ORF1 specific primer (ε and ζ), ORF2 primer (η and θ) and scFv-Fc specific primer (γ and δ) were designed for detecting gene delivery.

The results shown in Figure 7 indicate that ORF2 amplified bands were observed in the genomes of most (17 out of 18) clones. However, ORF1 was only detected in three clones. The neomycin resistance gene and scFv-Fc gene delivered by the LINE-1 vector could be identified in all clones.

This result indicated that synthetic LINE-1 vector is an efficient gene delivery

method.

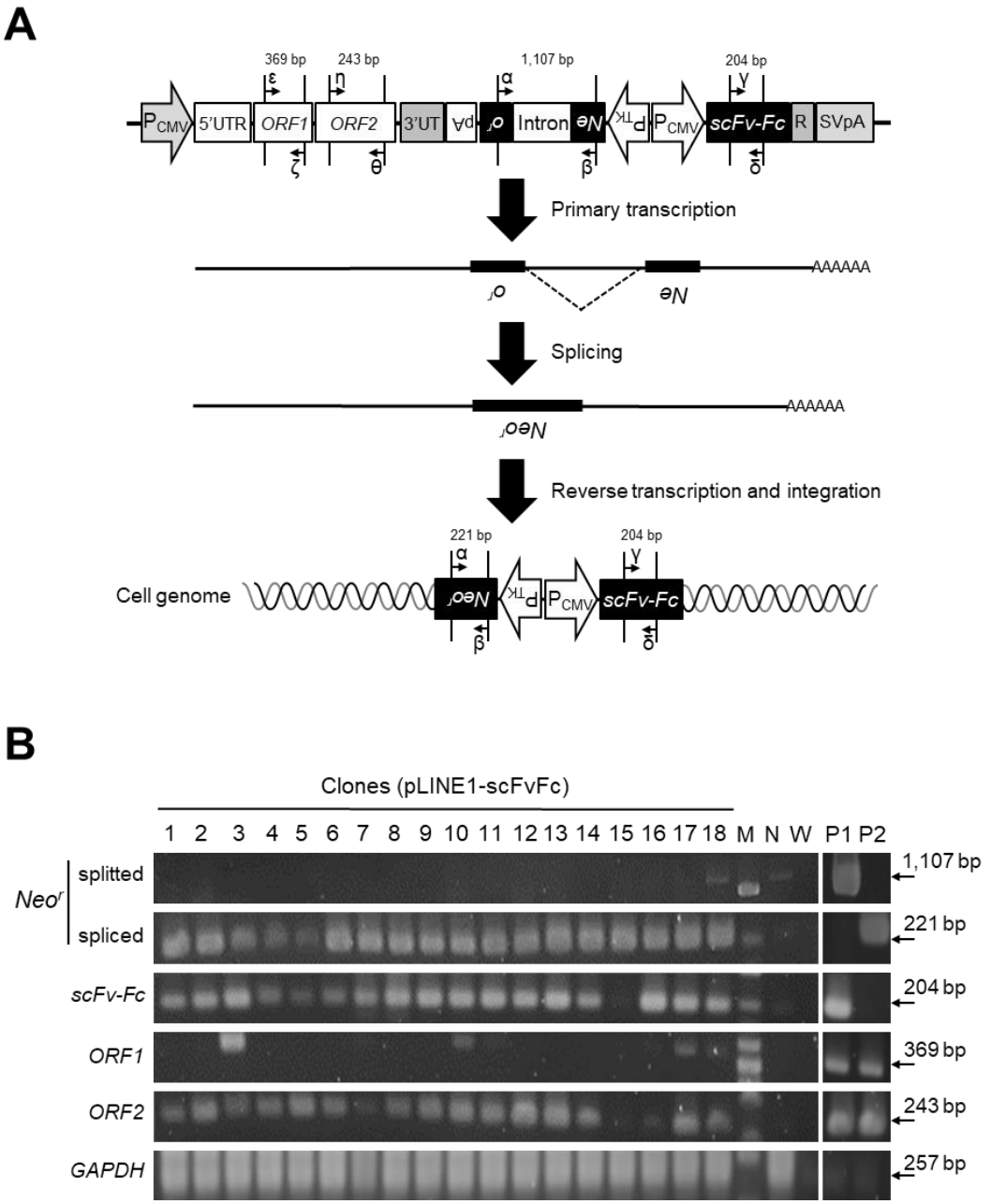


Figure 7. Recombinant CHO cells established by LINE-1 based gene delivery tool. (A) Schematic drawing of retrotransposition assay (B) Genomic PCR for detecting the integration of fragment in LINE-1 and the control (GAPDH). Lanes 1–18, engineered CHO cells generated using LINE-1 derived gene transfer tool; M, HincII-digested ΦX174 molecular weight marker; N, distilled water; C, original CHO-K1 genome; P1, Neor expression vector (pIRES-DsRedExpress); P2, pLINE1-scFvFc vector containing split Neor.

Retrotransposition using an ORF2-separated “feeding” system

Integration of the full-length LINE-1 vector expressing ORFs into the genome can lead to unexpected retrotransposition, which can lead to genome instability. Therefore, a regulated LINE-1 vector that can terminate the retrotransposition after integration was developed.

In the sequence of LINE-1, ORF1 and ORF2 are sandwiched between two UTRs. ORF1 and ORF2 are essential proteins for retrotransposition. The ORF2 protein has endonuclease activity that cleaves the genomic DNA at the new insertion site and reverse transcriptase activity that creates a DNA copy from LINE-1 RNA. ORF2 catalyze the integration of LINE-1 into the genome. On the other hand, the ORF1 protein is known to play a role of inducing the combination of ORF2 protein and LINE-1 RNA, although its detailed function has not been elucidated.

Therefore, a ORF proteins separated providing system for controlling the retrotransposition since ORFs plays an important role for retrotransposition was developed (Figure 8).

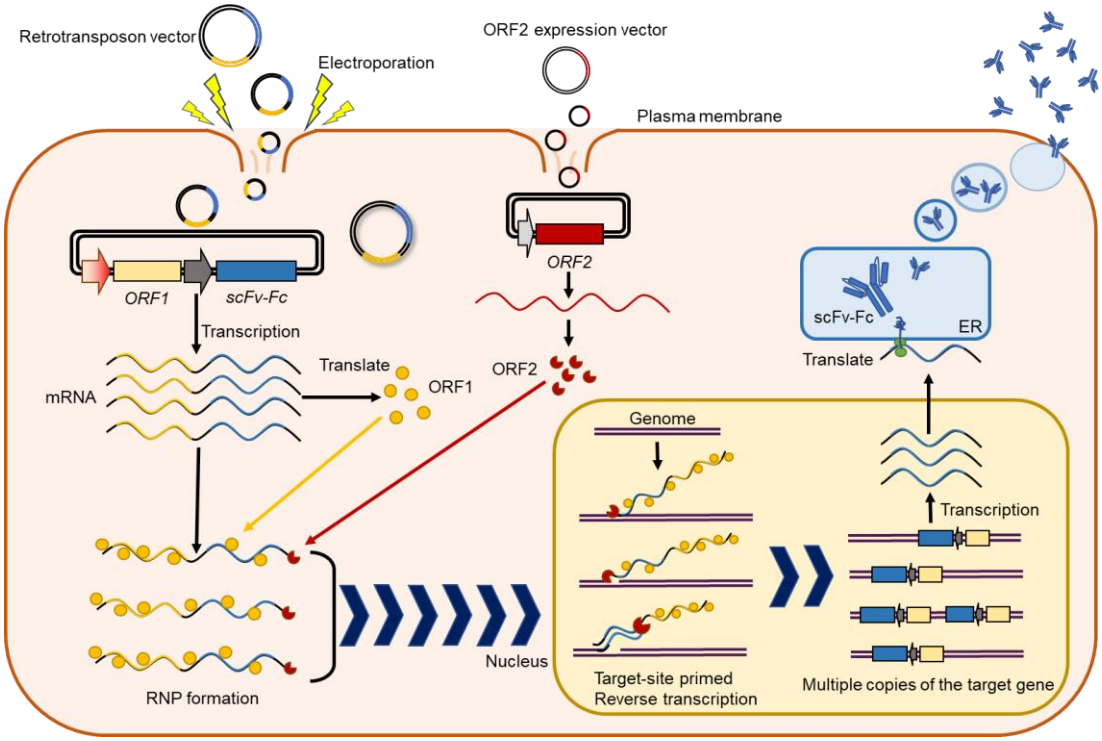


Figure 8. Schematic representation of LINE-1 gene delivery method.

To regulate the retrotransposition events, ORF proteins separated providing systems were designed. ORFs removed LINE-1 vector including ORF1 removed LINE-1 vector (pLINE1ΔORF2-scFvFc), ORF1-2 removed LINE-1 vector(pLINE1ΔORF1-2-scFvFc) and ORF1 mutated ORF1 removed vector (pLINE1mORF1-ΔORF2-scFvFc) was constructed. ORFs expression vectors including ORF2 expression vector(pORF2) and ORF1-2 expression vector (ORF1-2) were constructed. pLINE1ΔORF2-scFvFc was co-transfected with pORF2 or pORF1-2. After G418 selection, no drug resistance colony was formed with only ORFs removed LINE-1 vector. Cells co-transfected with ORF2 removed LINE-1 vector and ORFs providing vectors, on the other hand, formed drug resistance colonies. Clones were established and genomic PCR was performed for evaluating the gene delivery. Neomycin resistance gene and scFv-Fc gene amplicons were detected in most of clones established by ORFs providing system. The ORF1-2 providing system using pORF1-2 and pLINE1ΔORF2-scFvFc showed ORF1 detected in 7 out of 10 clones and ORF2 detected in only 3 out of 10 clones. On the other hand, the ORF2 providing system using pORF2 and pLINE-1ΔORF2-scFvFc showed a higher detection of ORF1 in most clones (13/16), but very few clones (2/16) showed detection of ORF2. This result indicated that ORF2 removed LINE-1 vector could mediate a ORF2 protein inducible retrotransposition. Therefore, the retrotransposition events with ORF2 removed LINE-1 vector could be regulated (Figure 9).

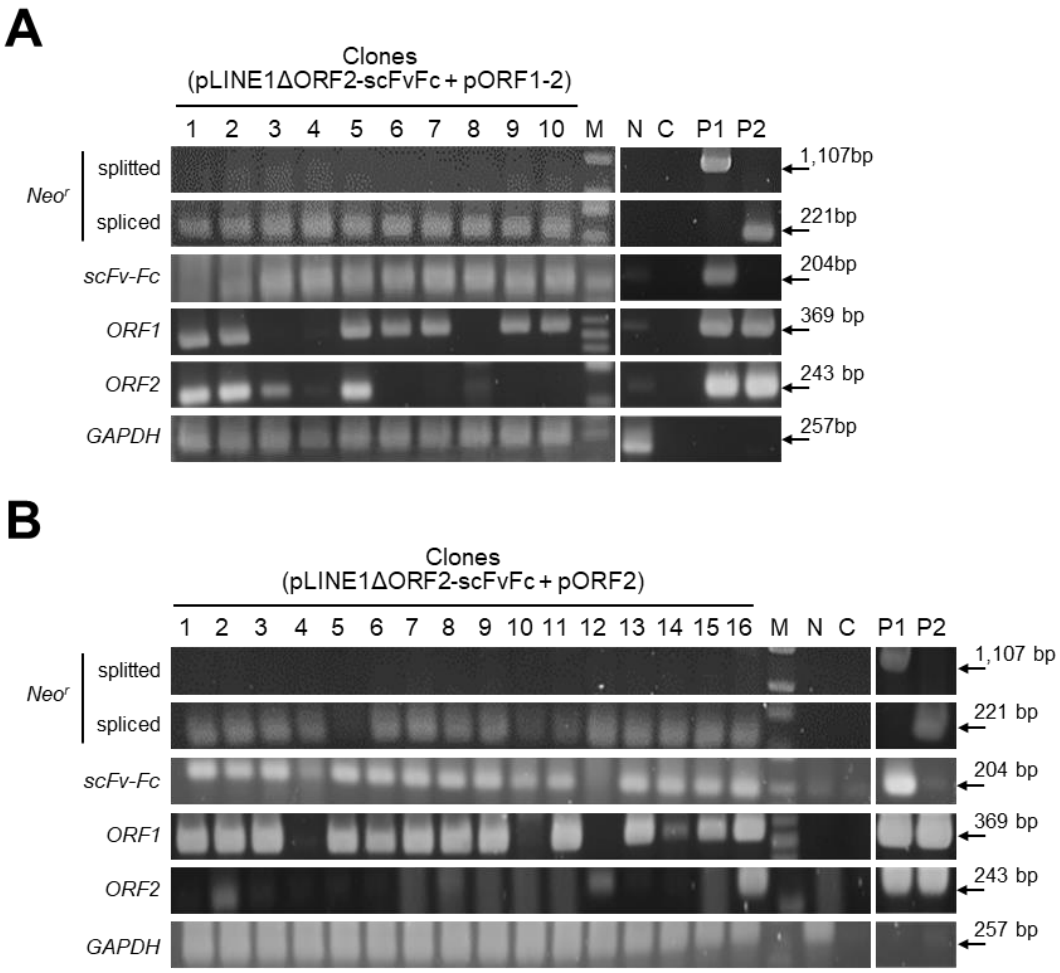


Figure 9. Retrotransposition assay of CHO clones generated using the ORFs separately providing LINE-1 vector. (A) Genome PCR amplicon of recombinant CHO cells generated by ORF1-2 providing system; (B) Genome PCR amplicon of recombinant CHO cells generated by ORF2 providing system

Retrotransposition efficiency

Gene integration efficiency is a crucial aspect in designing gene delivery strategies. In cell genome modification mediated by gene editing technology, the efficiency of NHEJ-mediated gene knock-out is much higher than that of HDR-mediated gene integration after double-strand break induction. Thus, Suzuki et. al have develop a homology-independent targeted insertion (HITI) method that can achieve targeted gene integration with NHEJ for enhancing the knock-in efficiency. Although the knock-in accuracy is not as well as HDR, HITI technology is still widely used to construct engineered cells, indicating that gene integration efficiency is important for genome modified cells generation.

To evaluate the efficiency of retrotransposon-based gene delivery method, three delivery system including the ORF2 providing system (pLINE1ΔORF2-scFvFc and pORF2), ORF1-2 providing system (pLINE1ΔORF2-scFvFc and pORF1–2) and intact type (pLINE1-scFvFc) were transfected into CHO cells. 20,000 transfected cells were seeded in each well for evaluating the colony number. The colonies after drug screening were counted to be 26.3, 23.3 and 36.6, respectively. The deleting of ORF-2 coding region did not significantly influence gene delivery efficiency if ORF-2 is separately provided. The gene delivery efficiency of ORF2 providing system, ORF1-2 providing system and intact type LINE-1 gene delivery system was almost same. ORF1 and ORF2 removed LINE-1 vector (pLINE1ΔORF1–2-scFvFc), mutated ORF1 LINE-1 vector (pLINE1mORF1-ΔORF2-scFvFc) formed no drug resistant colonies or formed few colonies even though ORF1 and ORF2 were provided by pORF1–2 or pORF1 + pORF2. These results indicated that not only ORF1 protein, but also ORF sequence is essential for LINE-1 retrotransposition (Figure 10).

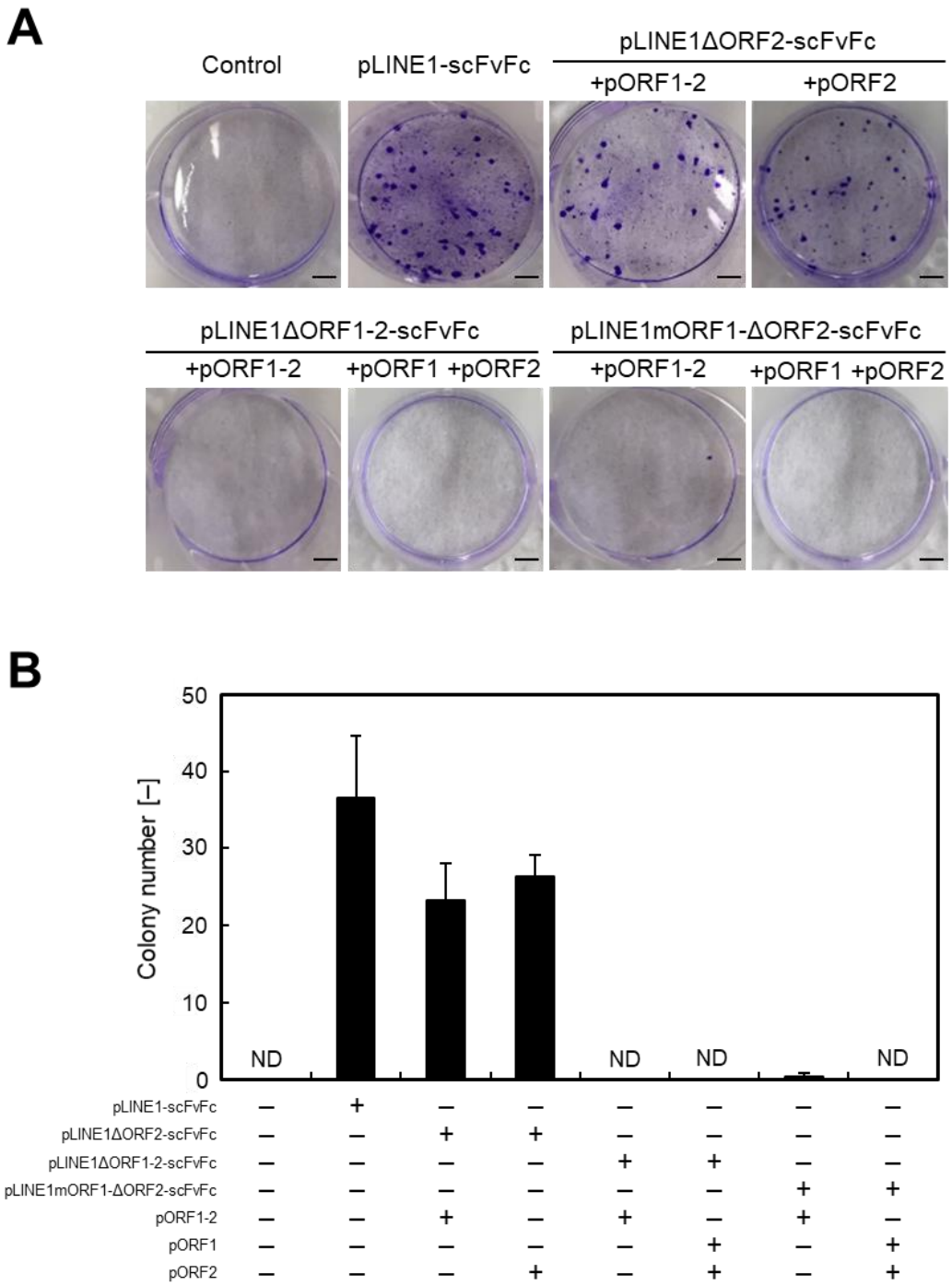


Figure 10. Retrotransposition assay for evaluating gene transfer efficiency using LINE-1 derived vectors. (A) Representative images of crystal violet-stained recombinant CHO colonies after G418 screening. Non-transfected cells were prepared as a control. Scale bars, 5 mm. (B) Number of G418-resistant recombinant CHO-K1 colonies generated using LINE-1 derived vectors.

Reverse transcription is thought to play an important role for LINE-1 mediated gene delivery. Thus, compounds that have the potential to improve the efficiency of LINE-1 gene delivery have been screened from database. It has been reported that 2–4 mM of caffeine could enhance the titer of lentivirus when added into culture medium. Caffeine may enhance the viral vector titter by repressing the DNA double-strand breaks repairing.

To enhance the gene delivery efficiency of LINE-1 transposon, three concentration of caffeine (0.04, 0.4 and 4 mM) were added into culture medium of LINE-1 transfected cells. The delivery efficiency was evaluated by retrotranscription assay. Efficiency of gene delivery method using LINE-1 vector was significantly enhanced for up to 1.9-fold with the addition of 0.4mM caffeine. With this strategy, gene delivery efficiency was evaluated to be up to 0.37% (Figure 11).

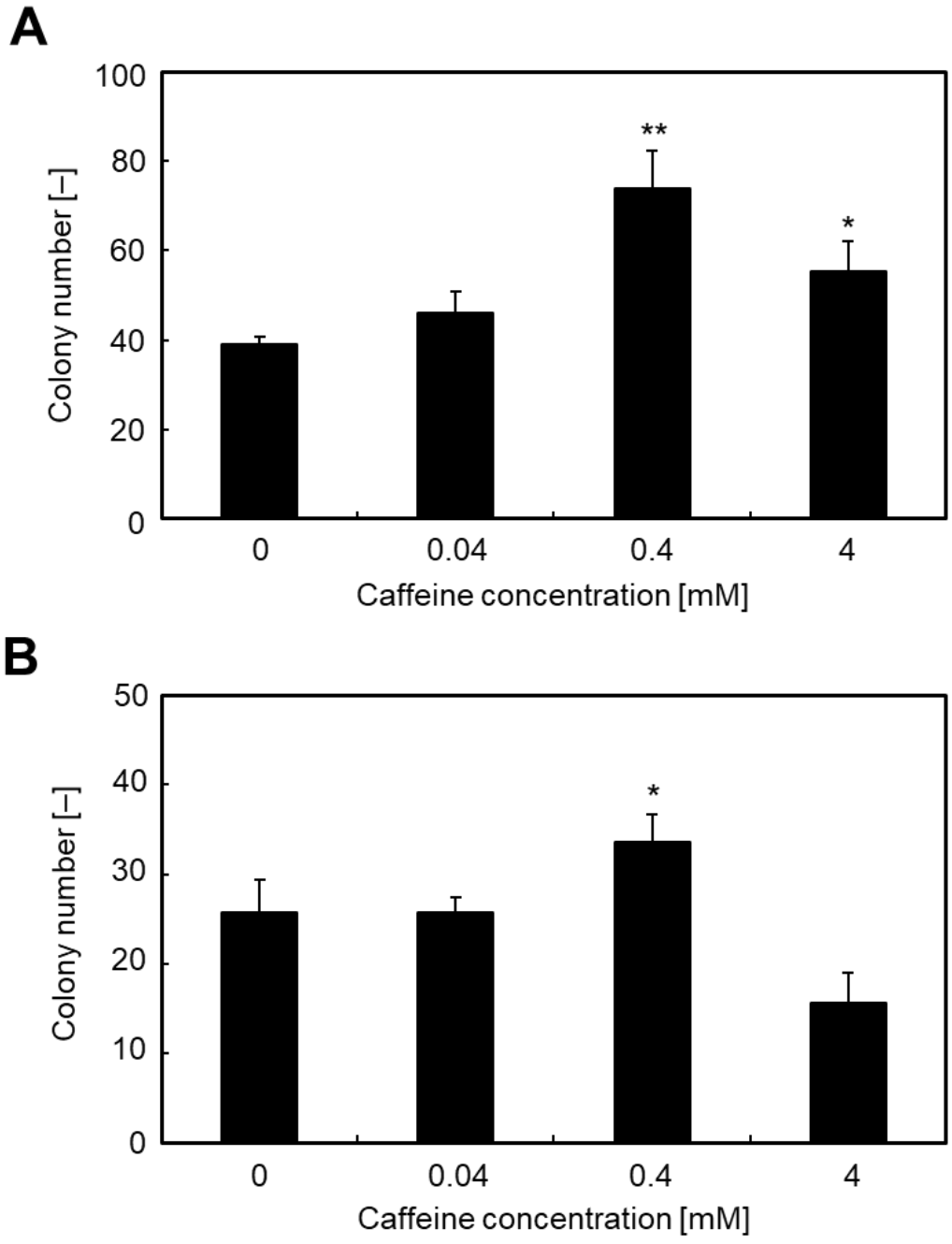


Figure 11. Colony-forming assay to evaluate retrotransposition efficiency when using the LINE-1 vectors with caffeine addtions. (A) Intact LINE-1 vector (pLINE1-scFvFc). (B) ORF2 feeding system (pLINE1ΔORF2-scFvFc + pORF2). Data represent the mean ± SD, n = 3. *P<0.05 and **P<0.01 compared with the absence of caffeine (Student's t-test).

Transgene copy number and scFv-Fc productivity of cell clones

To evaluate the ability of integrating biopharmaceutical genes using LINE-1 based gene delivery method, the copy number of antibody gene were assessed in clones established by ORF2 providing system, ORF1-2 providing system and intact type LINE-1 delivery system using quantitative real-time PCR. The copy number of scFv-Fc occurred to be ranged from 1 to 6 in 10 clones established by intact type LINE-1 vector. More than three copies of biopharmaceutical gene were confirmed in four clones. However, clones established by ORF2 providing system and ORF1-2 providing system was resulted to be containing 1 to 5 copies of the transgene.

Biopharmaceutical productivity was also evaluated in cell clones developed using ORF2 providing system, ORF1-2 providing system and intact type LINE-1 delivery system. Recombinant CHO cells were seeded at 24 wells in a cell density of 2.5×10^4 cells/well and cultured for 4 days. The scFv-Fc productivity was calculated using scFv-Fv concentration and cell number. The scFv-Fc productivity of best clones produced by ORF2 providing system, ORF1-2 providing system and intact type LINE-1 delivery system was exhibited 0.035, 0.121 and 0.134 pg per cell per day, respectively. The productivity of cells produced by ORF1-2 providing system and intact type LINE-1 delivery system was 3.5 and 3.8-fold higher than that of ORF2 providing system (Figure 12).

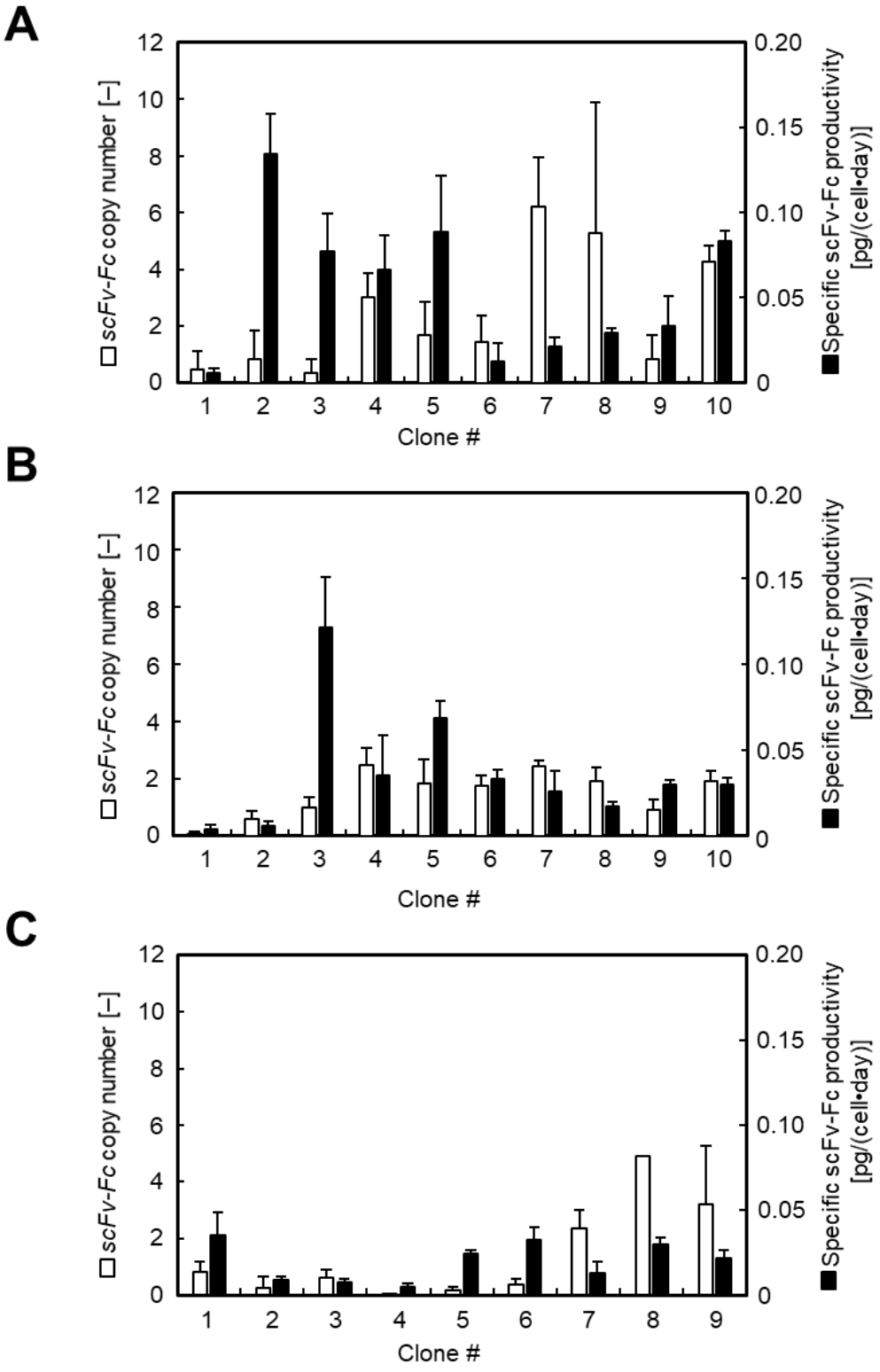


Figure 12. Copy number of the scFv-Fc gene in the genome and the scFv-Fc productivity of the genome modified CHO cells using LINE-1 derived gene delivery vectors. (A) Cells generated using intact type LINE-1 delivery system; (B) Cells generated using ORF1-2 providing system (C) Cells generated using ORF2 providing system.

2.4 Discussion

A gene delivery method based on LINE-1 vector was developed. However, only few clones were delivered full length LINE-1, which was consistent with previous studies that LINE-1 copies generated by retrotranscitption were usually truncated at 5'-end, which indicated that reverse transcription is terminated before completion. Nevertheless, the full length of additional sequence in 3'UTR was completely delivered, which was our destination.

A controllable retrotransposon vector using ORF2 removed LINE-1 retrotransposon have been developed. The truncations at the 5'-end were also detected in ORFs providing systems. However, more ORF1 amplicon were detected in ORFs providing system when compared with intact LINE-1 delivery system, which is possibly due to the shorter length of ORF2 removed LINE-1 vector.

A gene delivery efficiency of 0.37% using LINE-1 based gene transfer vector and 0.4 mM of caffeine in CHO cells is achieved. In contract, DNA transposon including PiggyBac have similar characteristics when compared with retrotransposon, both of them could delivery gene of interest in random site. It has been reported that the efficiency of gene delivery vector based on PiggyBac was observed to be about 0.027% in HEK 293 cells, which is lower than LINE-1 based gene delivery vector. Therefore, the LINE-1-based gene delivery tool was demonstrated to have the potential to be applied to construct genome-optimized engineered cells.

It has been reported that retrotansposition efficiency could be enhanced by compounds including heavy metals and methamphetamine. However, the toxicity of these drugs may affect the construction of engineered cells, which limited the application of these compounds on LINE-1 based gene transfer vector.

The copy numbers of cells established by intact LINE-1 delivery system is higher than of ORF2 providing system and ORF1-2 providing system. This result may be due to the lower transfection efficiency when co-transfecting two or more plasmid in ORFs providing system compared with transfecting one plasmid. Various copy number of scFv-Fc gene were observed in clones produced by LINE-1 based gene integrating tools.

Although it is expected that muti-copies of transgene were produced by “copy and paste” principle, the copy number is not enough when compared with PiggyBac transposon.

Recombinant protein gene integrated by LINE-1 gene delivery method were observed to be expressing antibody scFv-Fc. However, the productivity of antibody in engineered cells was not as high as cells produced by DNA transposon delivery tools or AGIS integration method, which may indicate that LINE-1 activity was strongly repressed by host cells. Coding sequence inside and nearby LINE-1 retrotransposon was reported to be undergo transcriptional silencing. Effort of overcome silencing including using high stable promoter such as EF1α may benefit the productivity. Nevertheless, scFv-Fc gene was integrated into CHO genome in a high efficiency and expressed in an acceptable level.

In summary, LINE-1 is a type of retrotransposon that widely exists in human genome, whose mechanism has been well investigated. However, LINE-1 has not be developed as an artificial synthetic biology element. In this chapter, a novel artificial gene element has been developed for gene delivery using LINE-1 retrotransposon. Transgene was detected in the engineered cells established using LINE-1 based transgene tool. Controllable LINE-1 vectors were engineered by separately providing ORF2 using an ORF2 removed LINE1 plasmid and a helper plasmid. The delivery of muti-copy recombinant protein gene were confirmed in genome of recombinant CHO-K1 cells. Recombinant protein production was confirmed in engineered cells generated by LINE-1 based gene delivery element. The LINE-1 based gene delivery system may serve as new artificial element for engineering genome modified cells.

Chapter 3

Development of an artificial transcriptional activation system for RNA detection

3.1 Introduction

Synthetic biology elements that regulate gene transcription can determine cell fate and function. Several transcriptional regulatory elements have been developed for the construction of gene circuits. However, the number and function of these elements are still limited. Therefore, in this study, a novel transcriptional regulation element for detecting RNA was developed.

Gene circuits is a type of intracellular machines that can sense intracellular or extracellular condition changes, process signals, and precisely regulate cellular functions by chemically synthesizing or assembling specific genetic devices, same as electronic circuit **[56]**.

Similar to cellular response to the change of environmental conditions such as temperature and oxygen concentration, RNA expression contains a wide range of information including physical conditions, chemical environments and endogenous signal **[33, 57]**. Therefore, RNA sensors triggered by defined RNA is a potential solution to address the challenge. In prokaryotes, efficient RNA responsive gene expression systems have been well developed by switching the secondary structure of ribosome-binding site with toehold switch **[45, 58, 59]**. The toehold switch strategy has achieved a 400-fold dynamic range, orthogonal regulation of 12 gene expression and basic logic calculation in bacteria. With similar principles, in eukaryotes, gRNA regulators and start codon regulators have been developed for detecting RNA based on toehold switch **[60-62]**. Although a variety of mammalian RNA sensors have been reported, the narrow dynamic range (~5-fold for Arbitrary RNA **[61, 62]** and ~50-fold for defined RNA**[63]**) and poor composability severely limits their applications due to the lake of ribosome-binding site.

Inducible complex formation strategy has been widely used as an efficient method for the study of protein-protein interactions **[64]** and the construction of synthetic elements **[65]**, in which two proteins essential for function (for example, DNA binding protein and transactivator) were separated and the triggers, including molecules **[66-68]**, lights **[35]** and protein-protein interactions **[69]**), induces the attaching of two parts and result in functioning. Sensor of vascular endothelial growth factor (VEGF) have been engineered using split-dCas9-VP64 and split-TEV. The binding of VEGFR-1 and VEGFR-2 induced by VEGF attach two parts of split-TEV and split-dCas9. The releasing of dCas9-VP64 result in a VEGF dependent 1,179-fold expression activation of target gene **[69]**. Thus, the inducible complex formation strategy provides a toolbox for designing synthetic elements with high efficiency and wide dynamic range.

RNA binding domain such as bacteriophage coat proteins **[70-72]** and RNA targeting Cas proteins **[2, 73]** have been widely used for the developing of live-cell RNA imaging system. Bacteriophage coat proteins can be combined with the RNA stem-loop structure in their genome. By fusing bacteriophage coat proteins with fluorescent proteins and introducing more than forty repeated of bacteriophage stem loops to 3'UTR of mRNA as a tag, subcellular localization of mRNA can be visualized by fluorescence microscopy. Some of RNA imaging system using RNA binding domains have single-molecule resolution **[74]**, suggesting that RNA binding domains are efficiency elements for detecting RNA. To date, however, there has been no report about developing high efficiency RNA sensor for inducible transcriptional regulation using RNA binding domains (Figure 13).

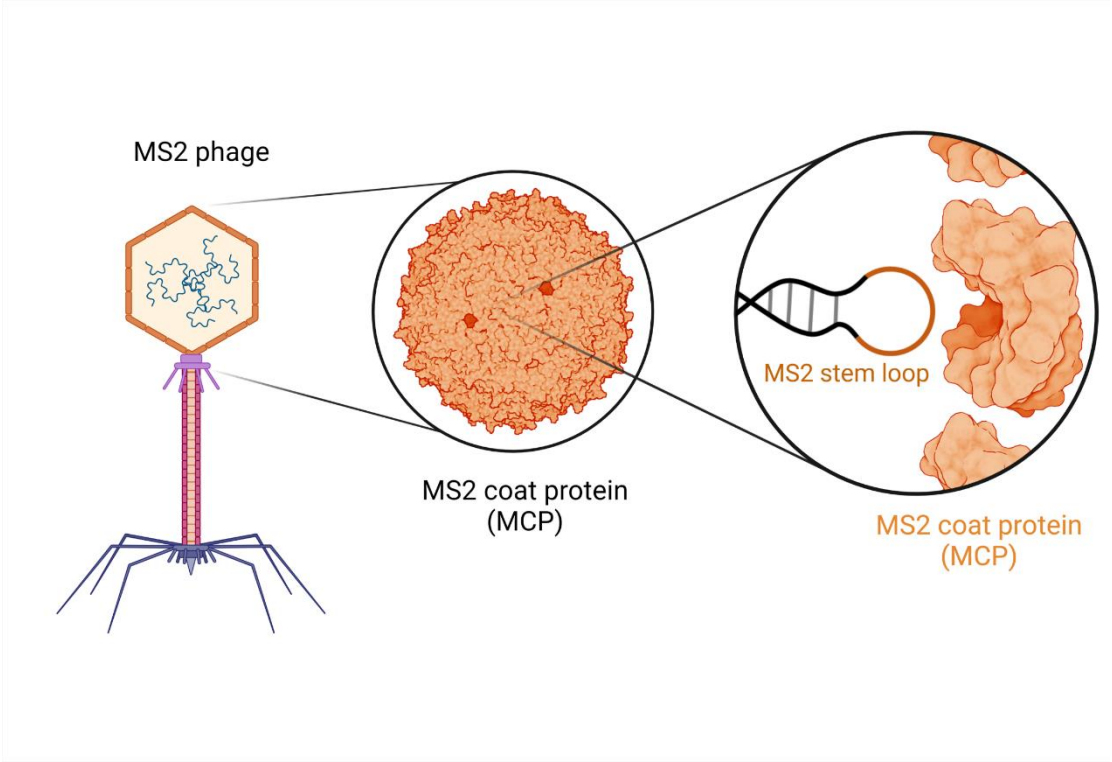


Figure 13. Diagram illustrating the binding of MS2 coat protein to the MS2 stem loop in the MS2 phage.

To overcome the low dynamic range of RNA detection in mammalian cells, in this study, a wide dynamic range RNA sensor named Specific Ribonucleic acid Responsive gene activation (SONAR) was developed based on inducible complex formation strategy. DNA binding domains, RNA binding domains and transcriptional regulation domains were organized as parts of RNA sensor. Two RNA binding domains were fused with DNA binding domain and transcriptional regulation domain, respectively. RNA recruiting both two RNA binding domains can be the trigger that attach DNA binding domain and trans-activator together for inducing gene expression. Here, fusion protein expression vectors were constructed for RNA sensing and tested the sensitivity for the detection of different types of RNA in mammalian cells. This study identifies that SONAR could be a high efficiency RNA sensor with wide dynamic range and provides guideline for future improvements (Figure 14).

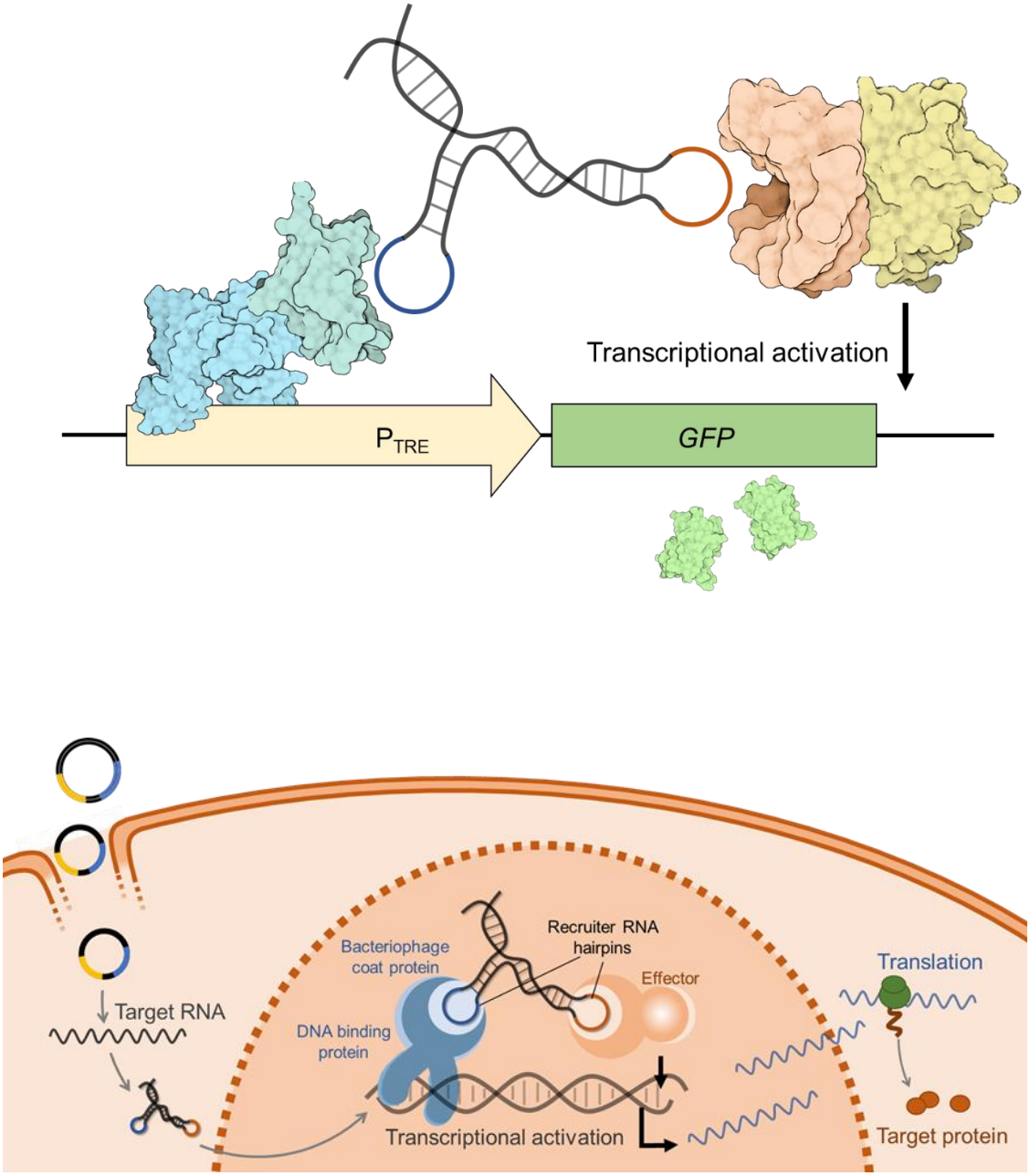


Figure 14. Schematic representation of a SONAR RNA detection system, based on phage coat protein (MCP and PCP), DNA binding domain (tetR), inducible promoter (PTRE), P65-HSF1 transcriptional activation domain (TA) and reporter gene (GFP). Without trigger RNA (MS2-PP7, trRNA).

3.2 Materials and methods

Cell culture

CHO-K1 cells (RIKEN, Tsukuba, Japan) were maintained in same condition described in Chapter 2.

Plasmid construction

For plasmid used for cell line establishment, pPB-UAS-GFP-ZeoR, pcDNA3.4-RFP and pPB-TRE-*tTA*-TRE-GFP-puroR were chemically synthesized. GFP reporter gene expression vector, pPB-TRE-GFP-puroR, were generated by remove *tTA*-TRE from pPB-TRE-*tTA*-TRE-GFP-puroR. RFP reporter gene expression vector, pPB-UAS-RFP-ZeoR were prepared by replacing GFP sequence in pPB-UAS-GFP-ZeoR with RFP sequence from pcDNA3.4-RFP.

For plasmid used for RNA sensing, pcDNA3.4(Thermo Fisher Scientific) was used as the backbone for most fusion protein expression vectors. Lenti MS2-P65-HSF1_Hygro was a gift from Feng Zhang (Addgene plasmid # 61426) **[75]**(Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex). MS2-P65-HSF1 was inserted into multiple cloning site of pcDNA3.4 to generate pCMV-MCP-TA. PCP-tetR, N22-Gal4 and Comp-tetR expression vector chemically synthesized by Thermo Fisher Scientific with pcDNA3.4 backbone (pCMV-PCP-tetR, pCMV-N22-Gal4, and pCMV-Comp-tetR). RNA binding domain digested with EcoRI and BamHI was ligated with DNA binding domain or trans-activator for exchanging RNA binding modules to generate pCMV-MCP-tetR, pCMV-PCP-TA and pCMV-N22-tetR. Expression vector of dCas9, pSLQ2814 pPB: CAG-Puro-WPRE PGK-VPR-tagBFP-SpdCas9 was a gift from Stanley Qi (Addgene plasmid # 84247). All in one RNA sensing vector (pCMV-TA-PGK-PCP-dCas9-U6-TRE-gRNA) was generated by replacing VPR sequence in pSLQ2814 with PCP sequence and inserting MCP-TA expression sequence and TRE-gRNA expressing sequence.

For plasmid used for trigger expressing, pU6-4X(MS2-PP7)-Sirius, pU6-4X(MS2-com)-Sirius, pU6-4X(BoxB-PP7)-Sirius, pU6-2X(MS2-MS2-PP7-PP7)-Sirius and pU6-4XMS2-4XPP7 were prepared by chemically synthesizing in accordance with previous reports **[76]** (CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging.). MS2-PP7, 2X(MS2-PP7), 4X(MS2-PP7), MS2-PP7-Sirius and 2X(MS2-PP7)-Sirius were polymerase chain reaction (PCR)-amplified from pU6-4X(MS2-PP7)-Sirius and inserted into backbone of U6 expression vector to generate pU6-MS2-PP7, pU6-2X(MS2-PP7), pU6-4X(MS2-PP7), pU6-MS2-PP7-Sirius and pU6-2X(MS2-PP7)-Sirius, respectively. To construct pEF1α-RFP-MS2-PP7 and pEF1α-MS2-PP7-RFP, 4X(MS2-PP7)-Sirius was PCR-amplified from pU6-4X(MS2-PP7)-Sirius and inserted into 3'UTR and 5'UTR (intron) of RFP mRNA. MS2-MS2-PP7-PP7 were PCR-amplified and toehold switch vector, pU6-TH-MS2-PP7, were prepared by golden gate assemble. EF1α promoter in pEF1α-RFP-MS2-PP7 were replaced with SV40 and minimum promoter prepared from PCR-amplification to generate pSV40-RFP-MS2-PP7 and pMin-RFP-MS2-PP7. RNA trigger with part of RFP sequence, pU6-RFP-trigger, were prepared by chemically synthesizing.

For plasmid used for cell line establishment, GFP reporter gene expression vector, pPB-TRE-GFP-puroR, were generated by remove *tTA*-TRE from pPB-TRE-*tTA*-TRE-GFP-puroR.

For plasmid used for RNA sensing, pcDNA3.4(Thermo Fisher Scientific) was used as the backbone for most fusion protein expression vectors. Lenti MS2-P65-HSF1_Hygro was a gift from Feng Zhang (Addgene plasmid # 61426) **[75]**. MS2-P65-HSF1 was inserted into multiple cloning site of pcDNA3.4 to generate pCMV-MCP-TA. PCP-tetR, N22-Gal4 and Comp-tetR expression vector chemically synthesized by Thermo Fisher Scientific with pcDNA3.4 backbone (pCMV-PCP-tetR, pCMV-N22-Gal4, and pCMV-Comp-tetR). RNA binding domain digested with EcoRI and BamHI was ligated with DNA binding domain or trans-activator for exchanging RNA binding modules to generate pCMV-MCP-tetR, pCMV-PCP-TA and pCMV-N22-tetR.

For plasmid used for trigger expressing, pU6-4X(MS2-PP7)-Sirius, pU6-4X(MS2-

com)-Sirius, pU6-4X(BoxB-PP7)-Sirius, pU6-2X(MS2-MS2-PP7-PP7)-Sirius and pU6-4XMS2-4XPP7 were prepared by chemically synthesizing in accordance with previous reports **[76]**. MS2-PP7, 2X(MS2-PP7), 4X(MS2-PP7), MS2-PP7-Sirius and 2X(MS2-PP7)-Sirius were polymerase chain reaction (PCR)-amplified from pU6-4X(MS2-PP7)-Sirius and inserted into backbone of U6 expression vector to generate pU6-MS2-PP7, pU6-2X(MS2-PP7), pU6-4X(MS2-PP7), pU6-MS2-PP7-Sirius and pU6-2X(MS2-PP7)-Sirius, respectively. To construct pEF1α-RFP-MS2-PP7 and pEF1α-MS2-PP7-RFP, 4X(MS2-PP7)-Sirius was PCR-amplified from pU6-4X(MS2-PP7)-Sirius and inserted into 3'UTR and 5'UTR (intron) of RFP mRNA. MS2-MS2-PP7-PP7 were PCR-amplified and toehold switch vector, pU6-TH-MS2-PP7, were prepared by golden gate assemble. EF1α promoter in pEF1α-RFP-MS2-PP7 were replaced with SV40 and minimum promoter prepared from PCR-amplification to generate pSV40-RFP-MS2-PP7 and pMin-RFP-MS2-PP7. RNA trigger with part of RFP sequence, pU6-RFP-trigger, were prepared by chemically synthesizing (Figure 15).

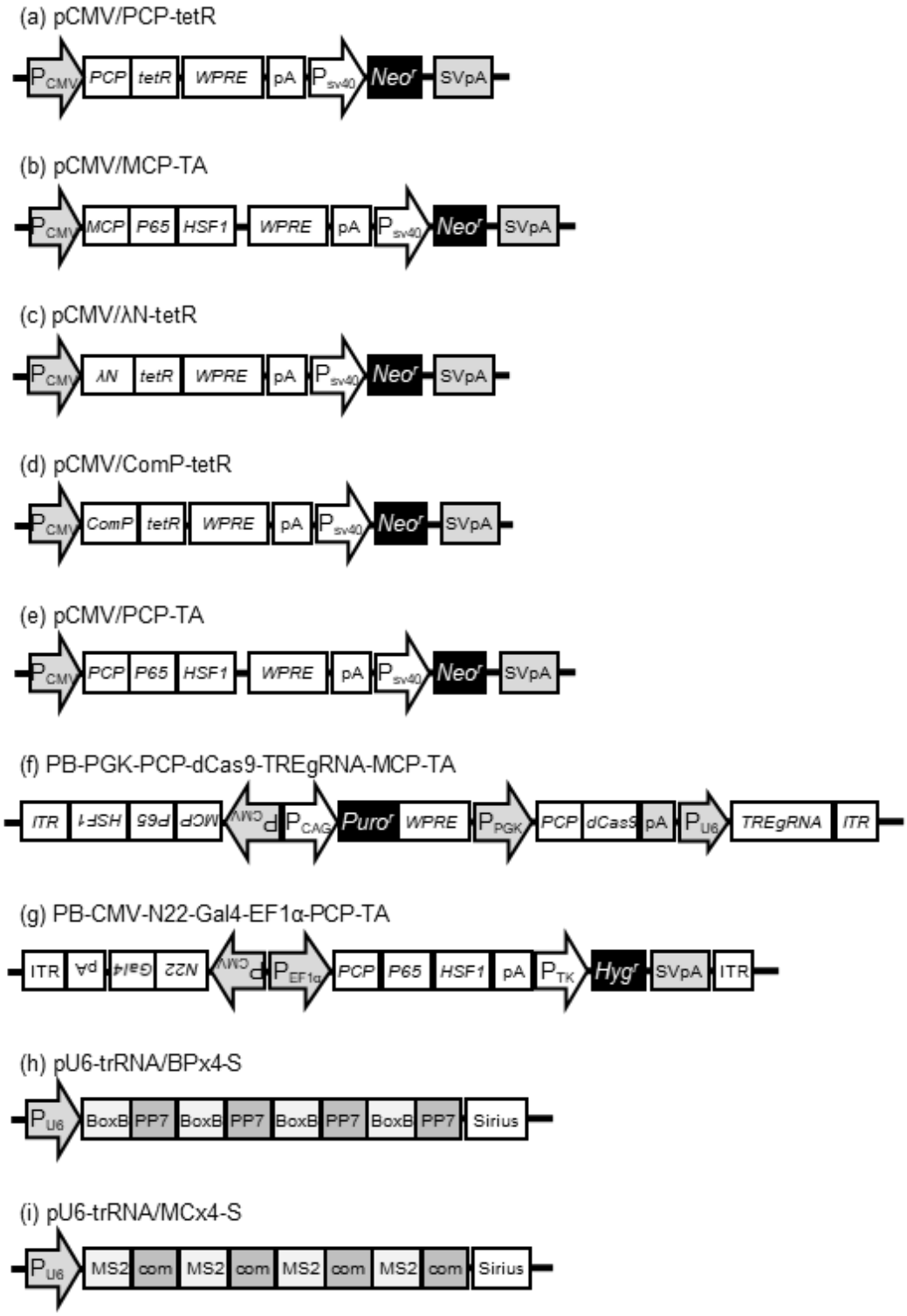


Figure 15. Schematic representation of plasmid vector for RNA sensing (a) CMV promoter-controlled PCP-tetR expression vector. (b) CMV promoter controlled MCP-TA expression vector. (c) CMV promoter controlled λN-tetR expression vector. (d) CMV promoter-controlled ComP-tetR expression vector. (e) CMV promoter-controlled PCP-TA expression vector.

RNA secondary structure prediction

The secondary structure of designed RNA was predicted using NUPACK web application and mfold web server **[77, 78]**. Toehold switches for arbitrary RNA detection were designed in silico using NUPACK nucleic acid sequence design package **[79]** and the interaction of switches and triggers were analyzed using NUPACK Analysis Algorithms **[80]**.

RNA inducible gene expression

CHO-K1 cells (5×10^4) were harvested from 100mm dish, Resuspended with Gibco™ Opti-MEM™ (Thermo Fisher Scientific) and transfected with totally 5000ng of plasmid vector from the NEPA21 Super Electroporator (Nepa Gene Co., Ltd, Chiba, Japan) in accordance with the manufacturer's protocol. The electroporation conditions for transfection were two poring pulses at 125 V for 10 ms with 50 ms pulse interval and 10% decay rate and 5 transfer pulses at 20 V for 50 ms with 50 ms pulse interval, 50% decay rate and polarity switching. The cells were seeded in 6-well plates with 2 mL of F-12 culture medium. For transient expression, the cells were seeded in three wells for each transfection condition and observed by fluorescence microscopy and cell sorter (Sony Biotechnology Inc, Tokyo, Japan) 48 hours after transfection.

TRE and UAS promoter controlling RFP and GFP expression cell line were established as follows. CHO-K1 cells were co-transfected with pPB-TRE-GFP-puroR and PBase expression vector for generating GFP expression cells. Five days after transfection, cells were transfected with *tTA* expression vector and cultured with 2 mL medium containing 50mg/L Puromycin from 24 hours after secondary transfection for selection. Seven days after Puromycin selection, clones were established, and the best clone was selected based on the on/off rate of GFP expression for generating CHO/TRE-GFP cell line. For establishing RFP reporter cells, CHO/TRE-GFP cell line were co-transfected with pPB-UAS-RFP-ZeoR and PBase expression vector and cultured with F-12 medium containing 100mg/L Zeocin (invivogen). The best clone of

RFP inducible expression was selected for preparing CHO/TRE-GPF/UAS-RFP cell line.

For evaluate the detection efficiency of mRNA, CHO/TRE-GPF were co-transfected with PBase expression vector and pEF1α-RFP-MS2-PP7, pSV40-RFP-MS2-PP7 or pMin-RFP-MS2-PP7. Cells were selected with 1000 mg/L Hygromycin B (invivogen) to generating RFP expression cells.

For transgene expression analysis in cells transient transfected with RNA sensor, GFP expression responsibility to specific RNA in cells was evaluated by fluorescence-activated cell sorting (FACS), real-time RT-PCR and microscope image analyzing. CHO reporter cells were analyzed 48 hours after transfection. Cumulated fluorescence intensity was the sum of fluorescence intensity of each cell read out from cell sorter and average fluorescence intensity was calculated by cumulated fluorescence intensity and cell numbers. Relative fluorescence intensity was calculated by dividing the average fluorescence intensity in transfected cells by that of untransfected cells. Phase contract image and fluorescent images were obtained under a BZ-9000 digital microscope (Keyence, Osaka, Japan). Image analyzing was performed, fluorescent intensity threshold was set based on cells without trigger and the fluorescent area ratio was calculated with fluorescent images using image analysis software (Keyence) by calculating the ratio of the cell area to fluorescent area as the GFP expression level.

For real-time RT-PCR analysis, total RNA from cells was extracted using a kit (RNAiso Plus; Takara Bio, Kusatsu, Japan) 48 hours after transient transfection, and RNA was reverse-transcribed into cDNA using reverse-transcriptase (RevaTra Ace; Toyobo, Osaka, Japan). The RNA level of GFP and RFP gene in transfected reporter cells was determined by real-time PCR using the PikoReal Real-Time PCR System (Thermo Fisher Scientific). Real-time PCR was performed using primers (Waltham, MA, USA), TaqMan probes and a quantitative PCR reagent (Thunderbird Probe qPCR Mix; Toyobo). Data analyzing was performed by PikoReal Software 2.0 (Thermo Fisher Scientific). Relative GFP or RFP expression was calculated by dividing mRNA level of transfected cells with CHO/TRE-GFP or CHO/TRE-GFP/UAS-RFP cells respectively.

3.3 Results

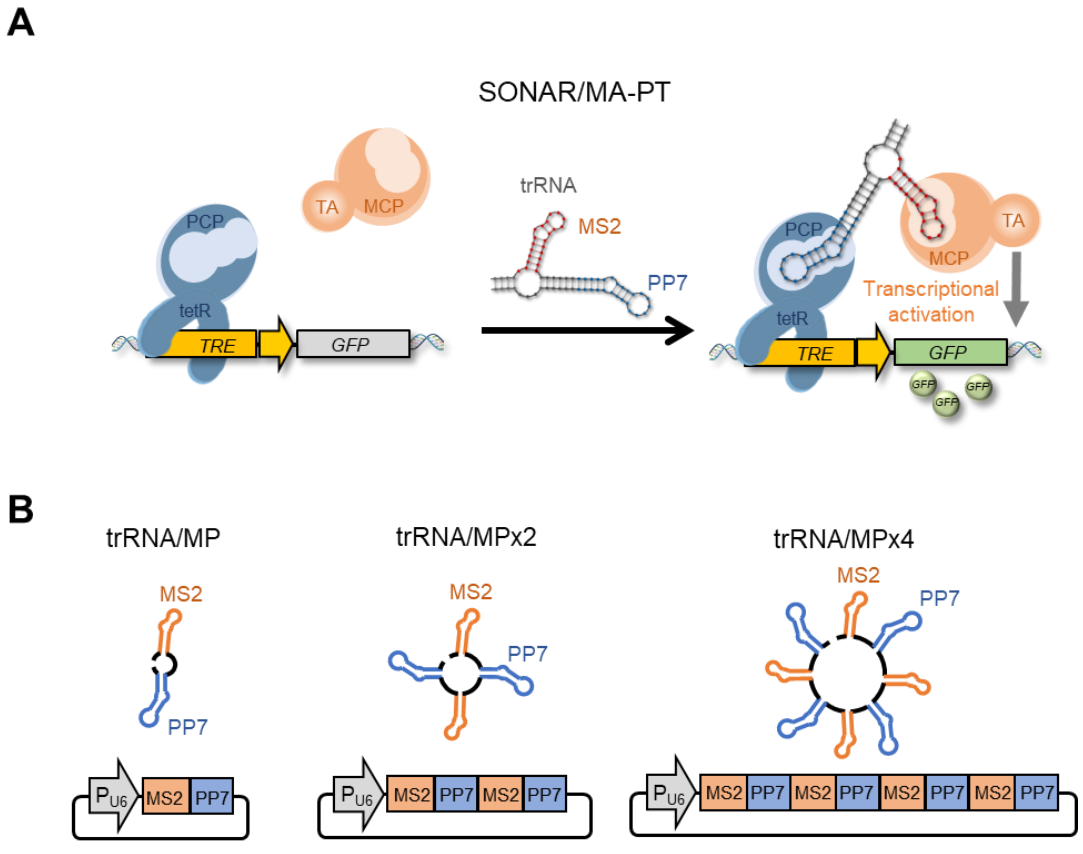
RNA responsive gene expression

To expand the gene circuit toolbox, a platform for detecting specific RNA was developed. In this study, an RNA responsive transgene expression regulation system was constructed using inducible complex formation strategy.

An RNA dependent trans-activation complex formation system using phage-derived coat proteins was designed. The well-characterized viral RNA sequences MS2 and PP7 aptamer hairpins were introduced into target RNAs, which are recognized by MS2 coated protein (MCP) and PP7 coated protein (PCP), respectively. Bacteriophage-derived coat proteins, MCP and PCP, were fused with P65-HSF1 transcriptional activation domain (TA) and DNA binding domain (tetracycline repressor protein, TetR) **[30, 81]**, respectively. RNA with both two types of RNA aptamer, MS2 and PP7, can be recognized by each coated protein to form nuclear ribonucleoprotein (RNP) complex containing MCP-TA, RNA and PCP-tetR. It is suspected that when cells transcribe target RNA, transcriptional activation complex can be formed and Cytomegalovirus (CMV) minimum promoter with tetR-responsive element (TRE) sequence (TRE/PCMVmin) can be activated. In cells without target RNA, on the other hand, the RNP cannot be assembled and TRE/PCMVmin has no activity. The system detecting MS2-PP7 containing RNA with MCP-TA and PCP-tetR was named SONAR/MA-PT. Above all, gene expression is designed to be regulated by the existence of specific RNA with this system.

To investigate the expression response of the RNA detection system, tetracycline responsive cell line (CHO/TRE-GFP) was established by PiggyBac transposon vector containing TRE/PCMVmin promoter controlled GFP reporter gene expression unit. Trigger RNA and Fusion proteins including PCP-tetR and MCP-TA were co-expressed in CHO/TRE-GFP cell line by transiently transfecting constitutive expression vectors to induce a trigger RNA responsive GFP expression. To adjust the valency of trans-activator recruited to the promoter, trigger RNA expression vector

with one, two, or four MS2-PP7 RNA aptamer (trRNA/MP, trRNA/MPx2 and trRNA/MPx4) were synthesized. Forty-eight hours after transfection, the cells were observed by phase contrast and fluorescence microscopies. The result, as shown in Figure 16, indicated that GFP was expressed in cells transfected with trigger RNA containing two or more ms2-pp7. To quantify GFP expression, fluorescence intensity was determined using fluorescence-activated cell sorting (FACS) and relative fluorescence intensity was calculated by dividing the average fluorescence intensity in transfected cells by that in CHO/TRE-GFP cells. The fluorescence intensity of cells with trRNA/MPx2 and trRNA/MPx4 existed 11-fold and 43-fold higher than that without target RNA, respectively (Figure 16). However, no statistically significant difference in GFP intensity between the cells expressed trRNA/MP and the cells without trigger RNA. These results indicated that cells transfected with fusion protein expression vectors could response to specific RNA with multivalent aptamers. Interestingly, cells only transfected with the fusion protein expression vectors exhibited lower GFP intensity (0.56-fold) than untransfected CHO/TRE-GFP cells.



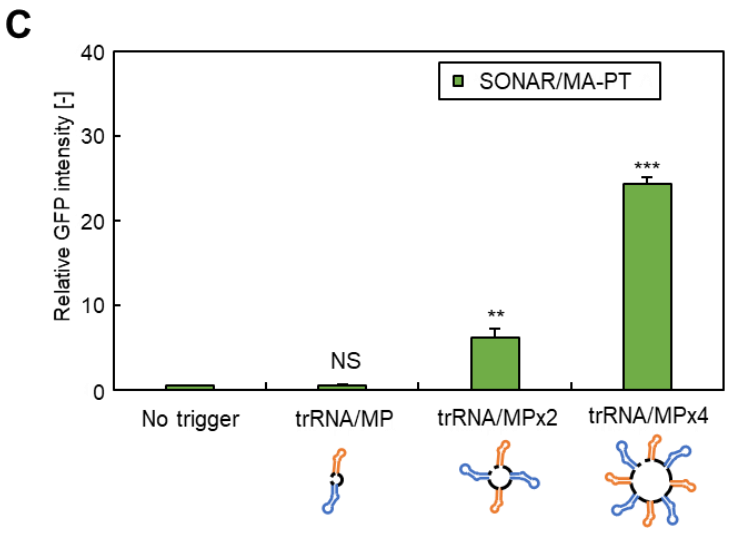


Figure 16. RNA inducible transgene expression using Specific Ribonucleic acid Responsive gene activation (SONAR) system (A) Schematic representation of a SONAR RNA detection system, SONAR/MA-PT, based on phage coat protein (MCP and PCP), DNA binding domain (tetR), inducible promoter (P_{TRE}), P65-HSF1 transcriptional activation domain (TA) and reporter gene (GFP). (B) Schematic illustration of trRNA and its expression vectors. trRNA/MP, trigger RNA with MS2-PP7; trRNA/MPx2, trigger RNA with two MS2-PP7; trRNA/MPx4, trigger RNA with four MS2-PP7. (C) Relative GFP intensity of the transfected CHO cells using SONAR/MA-PT with or without MS2-PP7 trigger RNA. MCP, MS2 coat protein; PCP, PP7 coat protein; tetR, tetracycline repressor; $P_{TRE,}$ minimal promoter with Tetracycline response element, GFP, copepoda Pontellina plumate green fluorescent protein; P_{U6}, human U6 promoter; MS2, stem loop derived from MS2 phage; PP7, stem loop derived from PP7 phage.

It was predicted that the RNA secondary structure obtained lower free energy after introducing the Sirius scaffold. Therefore, to optimize the RNA structure of RNA sensor SONAR, MS2-PP7-Sirius were generated by introducing Sirius scaffold to the 3' end of MS2-PP7 aptamer array (trRNA/MP-S, trRNA/MPx2-S and trRNA/MPx4-S) (Figure 17). Transient expression of PCP-tetR, MCP-TA and MS2-PP7-Sirius result in significantly increased GFP expression when compared with that without scaffold. GFP intensity induced by trRNA/MPx2-S and trRNA/MPx4-S was 70-fold and 210-fold higher than that without target RNA, respectively. Expression level of GFP induced by Sirius RNA aptamer was promoted approximately 5-fold when compared with RNA without Sirius scaffold. Thus, introducing Sirius to RNA aptamer array was valid for enhancing the sensitivity of SONAR RNA sensor.

The arrangement of RNA aptamers is one of possible factors that affect the efficiency of RNA sensor. RNA expression vectors with three types of MS2-PP7 aptamer arrangement, staggered type, spaced type and separated type, were chemical synthesized and co-transfected to reporter cells with fusion protein expression vector(Figure.17) (Staggered type: (MS2-PP7)x4-Sirius, trRNA/MPx4-S, two kinds of aptamers, ms2 and pp7 are placed alternately; Spaced type: (MS2-MS2-PP7-PP7)x2-Sirius, trRNA/MMPPx2-S,two sequential aptamers are composed as on unit, four times, (MS2-MS2-PP7-PP7)x2-Sirius; Separated type: MS2x4-PP7x4-Sirius, trRNA/Mx4-Px4-S, four sequential aptamers are one unit, two times.) Both types of aptamer arrangement were able to efficiently induce GFP expression (Figure.17). Interestingly, RNA array with stagged type and spaced type MS2-PP7 arrangement exhibited the highest and higher GFP intensity when compared with separated type, respectively.

To maximize the sensitivity of RNA sensor, RNA aptamer array used in following experiment was designed as four repeats and stagged type with Sirius scaffold unless mentioned.

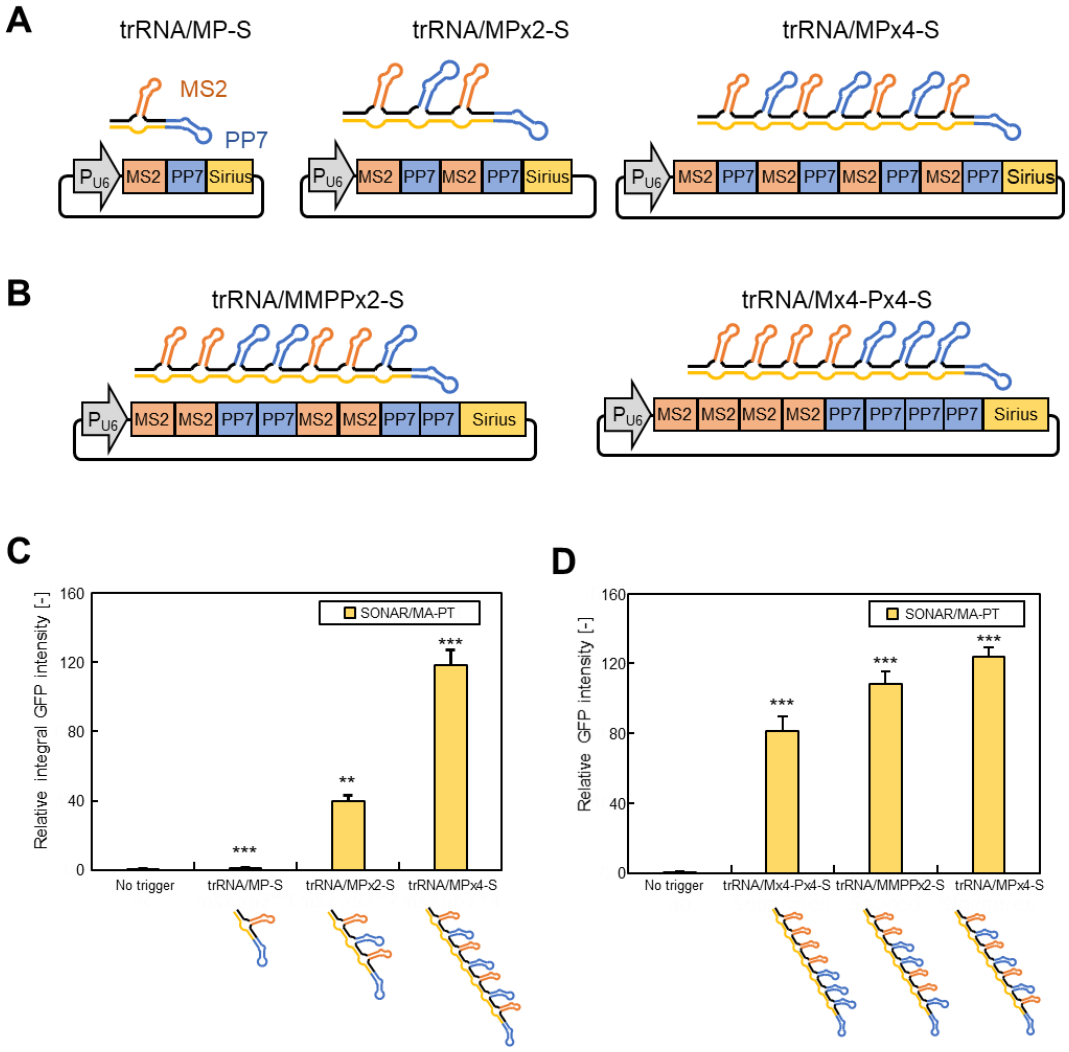


Figure 17. Optimization of trRNA structure in SONAR system (A) Schematic illustration of Sirius scaffold introduced trRNA and its expression vectors. trRNA/MP-S, trigger RNA with MS2-PP7 and Sirius scaffold; trRNA/MPx2-S, trigger RNA with two MS2-PP7 and Sirius scaffold; trRNA/MPx4-S, trigger RNA with four MS2-PP7 and Sirius scaffold. (B) Schematic illustration of trRNA with different aptamer arrangement and its expression vectors. trRNA/MMPPx2-S, trigger RNA with two MS2-MS2-PP7-PP7 and Sirius scaffold; trRNA/Mx4-Px4-S, trigger RNA with four MS2, four PP7, and Sirius scaffold. (C)Relative GFP intensity of the transfected CHO cells using SONAR/MA-PT and Sirius scaffold introduced trRNAs. (D)Relative GFP intensity of the transfected CHO cells using SONAR/MA-PT and trRNAs with different aptamer arrangement. Sirius, Sirius scaffold.

Messenger RNA inducible transcription activation

To evaluate the responsibility of RNA sensors to messenger RNAs, MS2-PP7 RNA array was introduced into either 3' UTR or 5' UTR of RFP to generate RFP constitutive expression vectors (pEF1α-RFP-MS2-PP7(trmRNA/R-MP) and pEF1α-MS2-PP7-RFP (trmRNA/MP-R), respectively) (Figure 18). To prevent interference of RFP expression, MS2-PP7 were inserted inside the intron of 5'UTR. Transient expression of aptamer containing RFP messenger RNAs, PCP-tetR and MCP-TA was induced in CHO/TRE-GFP cells and supposed to be resulted in a messenger RNAs responsive GFP expression. Extremely high intensity of RFP was detected in both pEF1α-RFP-MS2-PP7 and pEF1α-MS2-PP7-RFP transfected cells. Considering that strong red fluorescence will interfere the reading of green fluorescence in the flow cytometer, fluorescence microscope was applied for observing the cells with both red and green fluorescence. The percentage of the fluorescent area of the cells was calculated to quantitatively evaluate the expression of GFP or RFP. Messenger RNA induced GFP expression were confirmed but higher level of GFP expression was observed in pEF1α-RFP-MS2-PP7 transfected cells than that with pEF1α-MS2-PP7-RFP. However, fluorescence area induced by aptamers in messenger RNAs was smaller than that in small RNAs.

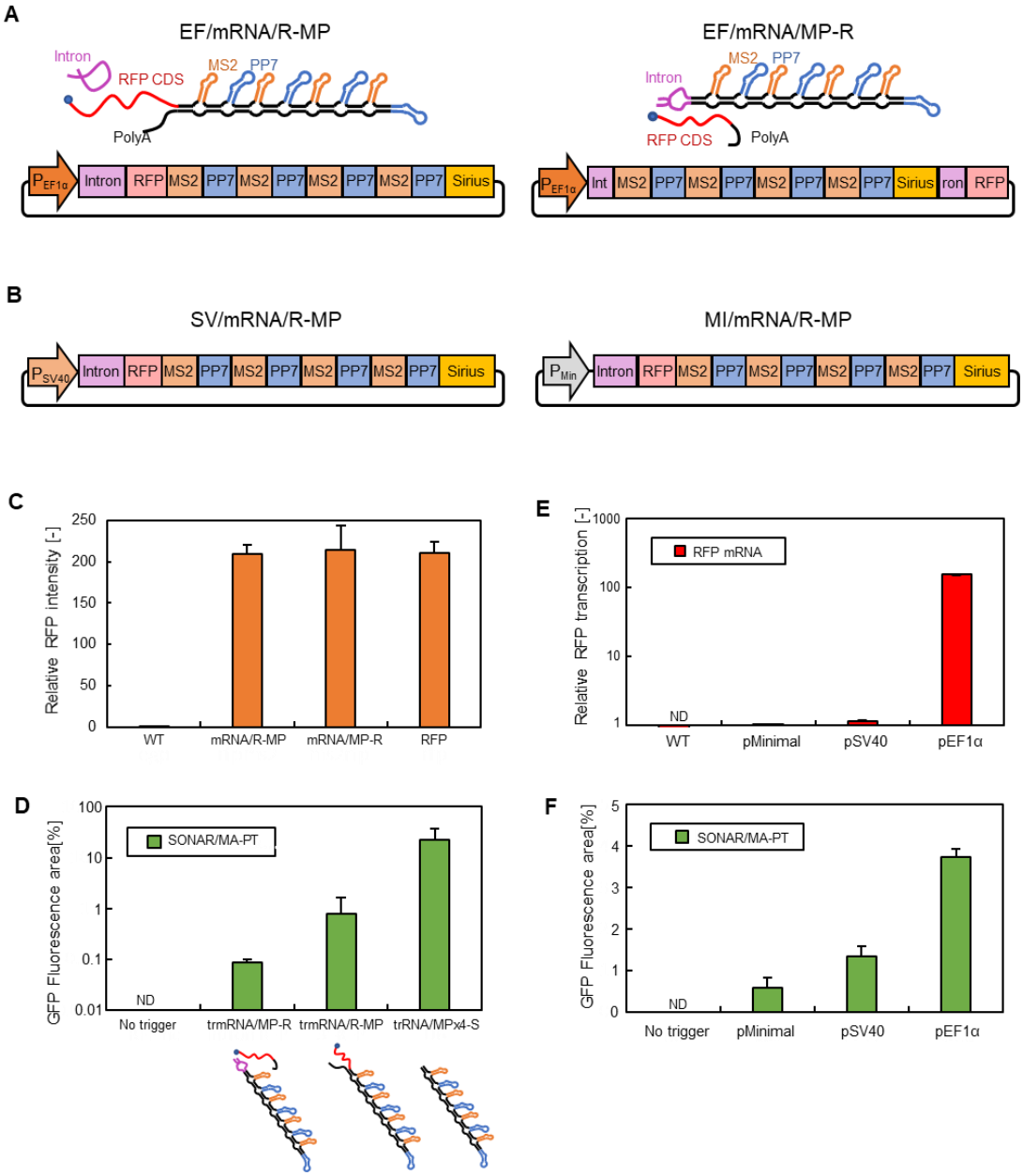


Figure.18 Messenger RNA inducible transgene expression (A) Schematic illustration of messenger trigger RNA and its expression vectors. mRNA/R-MP, RFP mRNA introduced with trRNA/MPx4-S in 3'UTR; mRNA/MP-R, RFP mRNA introduced with trRNA/MPx4-S in intron at 5'UTR. EF/mRNA/R-MP, EF1α promoter-controlled mRNA/R-MP expression vector; EF/mRNA/MP-R, EF1α promoter-controlled mRNA/MP-R expression vector. (B) Schematic illustration of messenger trigger RNA expression vectors with different promoter. SV/mRNA/R-MP, SV40 promoter-controlled mRNA/MP-R expression vector; MI/mRNA/R-MP, CMV minimal promoter-controlled mRNA/MP-R expression vector. (C) Relative RFP intensity of the transfected CHO cells using trigger mRNAs with different structures. No statistical-significant differences were observed in both two trigger mRNAs vs. RFP. (D) GFP Fluorescence area of the transfected CHO cells using SONAR/MA-PT and trigger mRNAs with different

structures. (E) Relative RFP mRNA level of the transfected CHO cells using RNA expression vectors with different promoter. (F) GFP Fluorescence area of the transfected CHO cells using SONAR/MA-PT and trigger mRNAs expression vectors with different promoter. $P_{EF1\alpha}$, chimpanzee Elongation Factor-1 alpha promoter;Intron, chimpanzee Elongation Factor-1 alpha intron; Int and ron, a split intron; RFP, green fluorescent protein; P_{SV40}, simian virus 40 promoter; P_{Min}, minimal cytomegalovirus promoter.

Efficient detection of RNA with aptamer tags was achieved. However, the detection of endogenous genes still relies on genome editing, which limits the application. RNA sensors response to arbitrary RNA is a possible solution for natural RNA detection.

To engineer natural RNA sequences responsive RNA sensor, RNA aptamer array was optimized, MS2-MS2-PP7-PP7-Sirius, with toehold switch. It has been demonstrated that exact nucleotide sequence in base-paired stem regions of MS2 and PP7 is unimportant but serves to maintain secondary structure for protein recruitment **[82, 83]**. To this end, a target RNA sensitive domain was introduced to 3' end of aptamer array as a blocking sequence, which is complementary to base-paired stem regions of PP7, resulting in the disrupting of PP7 secondary structure. Nineteen nucleotides toehold domain was attached to 3' end of aptamer array together with target RNA sensitive domain to generate strand displacement aptamer RNA (SDA RNA), both of blocking sequence and toehold are complementary to trigger RNA, parts of RFP gene. The binding of trigger RNA restores the stem-loop structure of PP7 via toehold-mediated strand displacement and induce gene expression. U6 promoter transcribed trigger sequence or CMV promoter transcribed RFP mRNA were transfected into CHO/TRE-GFP cells alongside TH-MS2-PP7, PCP-tetR and MCP-TA expression vectors. Small RNA trigger leaded to a 1.64-fold activation of GFP expression (Figure 19).

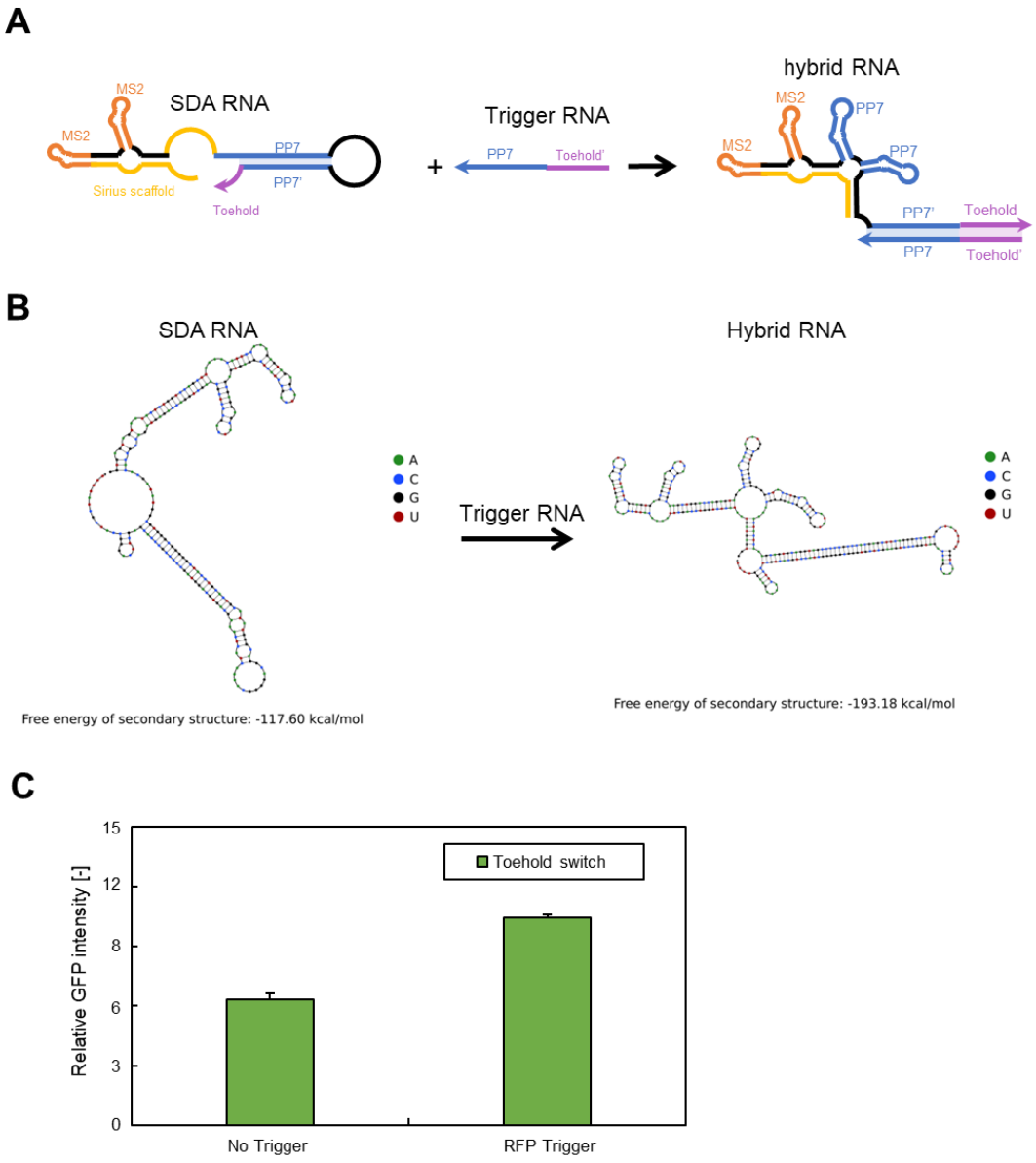


Figure 19. Arbitrary RNA inducible transgene expression using toehold switch and SONAR/MA-PT, (A) Schematic illustration of strand displacement reaction. (B) secondary structure prediction of SDA RNA and Hybrid RNA using NUPACK. (C) Relative GFP intensity of the transfected CHO cells using SONAR/MA-PT and SDA RNA with or without trigger RNA. Data represents the mean ± SD.

Characterization of RNA recruitment domains

To expand the detection range of RNAs, it was hypothesised that efficient RNA detection could still be achieved even if phage-derived coat proteins, MCP and PCP, were replaced with other RNA binding modules. The interchangeability between MCP

and PCP by swapping their position was tested. MCP and PCP were fused with tetR and TA, respectively. Although GFP expression could be detected when trigger RNA expression vector was co-transfected with the MCP-tetR/PCP-TA expression vectors, the expression level of reporter gene is not as high as that of PCP-tetR and MCP-TA. In addition to MCP/MS2 and PCP/PP7, bacteriophage coat protein, bacteriophage λ derived 22-amino acid peptide (λN) **[84]** and bacteriophage Mu mom-encoded C protein (ComP) **[85]**, were used as RNA recruiting modules, which recognize BoxB and com RNA sequences, respectively. Fusion protein λN-tetR expression vector was chemical synthesized and transient expressed with PCP-TA for detecting RNA containing BoxB-PP7 aptamer array(trRNA/BPx4-S). The system detecting BoxB-PP7containing RNA with PCP-TA and λN-tetR was named SONAR/PA-λT. RNA-dependent gene expression has been confirmed with λN-tetR/PCP-TA fusion protein. It has been reported that ComP recognize single-stranded RNA at the base of the com hairpin and thus attempts to construct com RNA aptamer array with stable secondary structure were failed. Here, base of the com hairpin structure was designed as a single-stranded 4nt loop and a 6bp double-stranded stem was introduced to stabilize the loop structure prevent mispairing. Due to the excellent thermal stability of the Sirius scaffold and the introduction of an additional stem structure, the base of com RNA hairpin was predicted to be stably exposed as a single-stranded structure even if double-stranded linkers were used to construct the aptamer array (Figure 20). The MS2-com aptamer array(trRNA/MCx4-S) induced the expression of reporter genes with fusion proteins, ComP-tetR and MCP-TA, called SONAR/MA-CT. Although both SONAR/PA-λT and SONAR/MA-CT can generate RNA responsible gene activation, RNA sensors with exchanged RNA binding modules exhibited lower transcriptional activation level when compared with MS2-PP7 RNA sensors, SONAR/MA-PT. The results indicates that although the potency of recruitment may affect the efficiency, RNA recruitment domains are interchangeable for RNA detecting

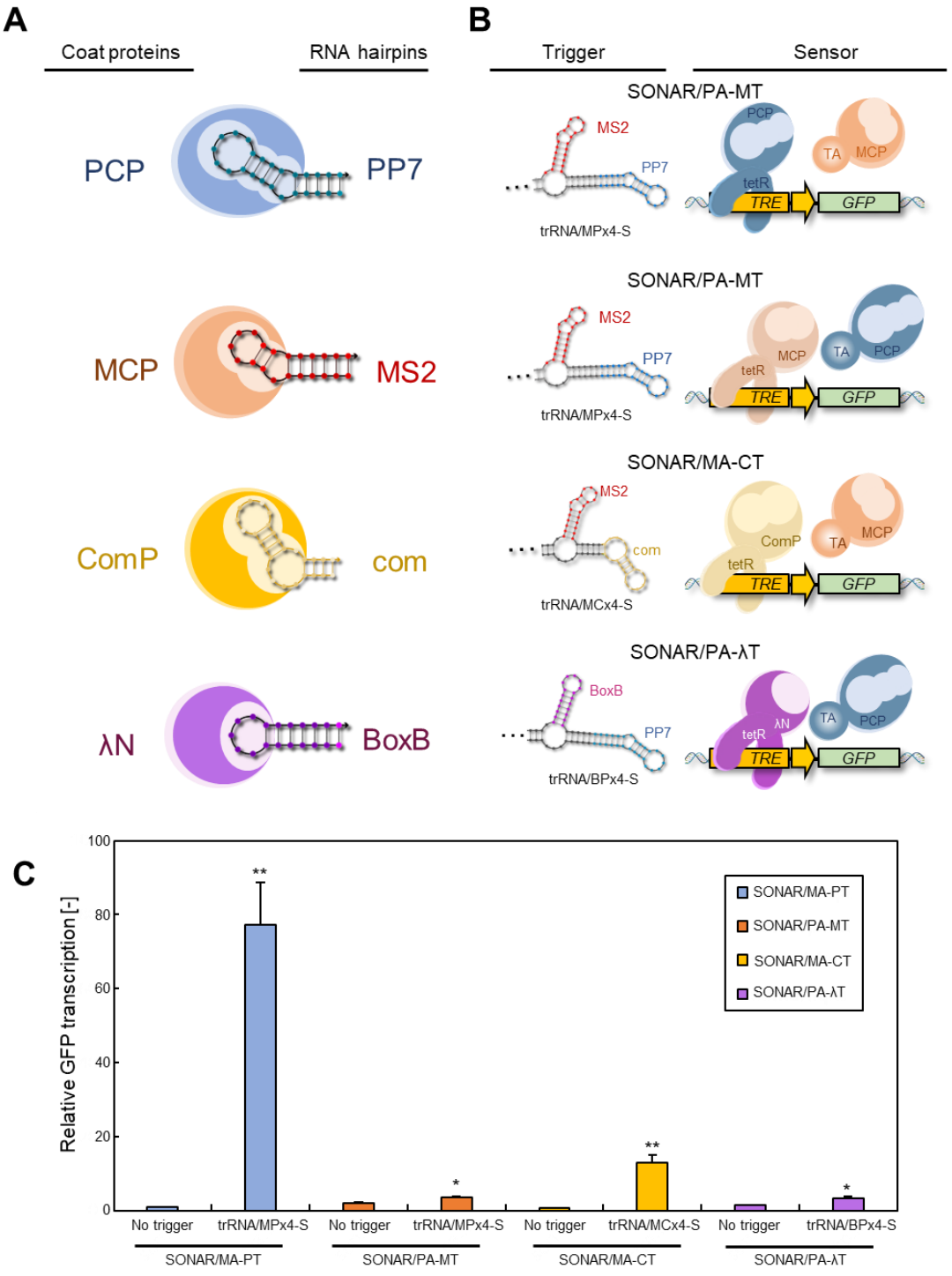


Figure 20. RNA inducible transgene expression using different RNA binding domain. (A) Schematic illustration of RNA binding domain and its binding element. PCP, MCP, ComP and λN can recruit PP7, MS2, com and BoxB, respectively. (B) Schematic representation of SONAR systems. SONAR/MA-PT, SONAR system using fusion protein of PCP-tetR and MCP-TA. (C) Relative GFP intensity of the transfected CHO cells using SONAR/MA-PT, SONAR/PA-MT, SONAR/MA-CT, SONAR/PA-λT and trRNAs. ComP, phage Com protein derived from bacteriophage Mu; com, a stem-loop RNA element derived from

bacteriophage Mu mom gene; λN, a 22-amino acid peptide derived from bacteriophage λ; BoxB, a stem-loop RNA element derived from bacteriophage λ.

Characterization of DNA binding domains

RNA responsive activation of TRE/PCMVmin have been achieved as described previously with tetR fusion protein. To develop a platform for activating different promoters with RNA sensor, tetR domain was replaced with other DNA-binding modules. The yeast gal4 regulatory protein (Gal4) have been reported to activate target gene by binding to upstream activating sequence (UAS) **[60]** .The Gal4 DNA binding domain was fused with RNA recruit domain λN to establish λN-Gal4/PCP-TA expression regulation system (SONAR/PT-λG). To establish UAS/TRE reporter cell line (CHO/TRE-GFP/UAS-RFP), the RFP gene expression unit controlled by the UAS and CMV minimal promoters (UAS-RFP) was chemically synthesized, which was integrated into the genome of CHO/TRE-GFP cells. Trigger RNA/BPx4-S inducible RFP expression in CHO/TRE-GFP/UAS-RFP cell line was confirmed with SONAR/PT-λG. Programmable DNA binding domain have been reported as a module for constructing artificial transcriptional activators for regulating endogenous gene expression. Among them, Endonuclease de-activated Cas9 (dead-Cas9, dCas9) has been developed as a modular DNA binding domain for targeted transcriptional regulation, which can bind to specific DNA sequences with programmable guide RNA. PCP was fused with dCas9 and TRE targeting gRNA was designed. To increase transfection efficiency, PCP-dCas9, MCP-TA and gRNA expression units were assembled into an all-in-one expression vector (PCP-dCas9-gRNA-MCP-TA) to generate dCas9 based RNA sensor, SONAR/MA-Pd. An trRNA/MPx4-S induced GFP expression in CHO/TRE-GFP cells after transiently transfected with SONAR/MA-Pd and trRNA/MPx4-S observed (Figure 21). However, the on/off ration was not as high as SONAR/MA-PT with trRNA/MPx4-S. Taken together, these results suggest that the RNA inducible gene expression system is a modular system in which DNA binding domain can be modified for a variety of purposes.

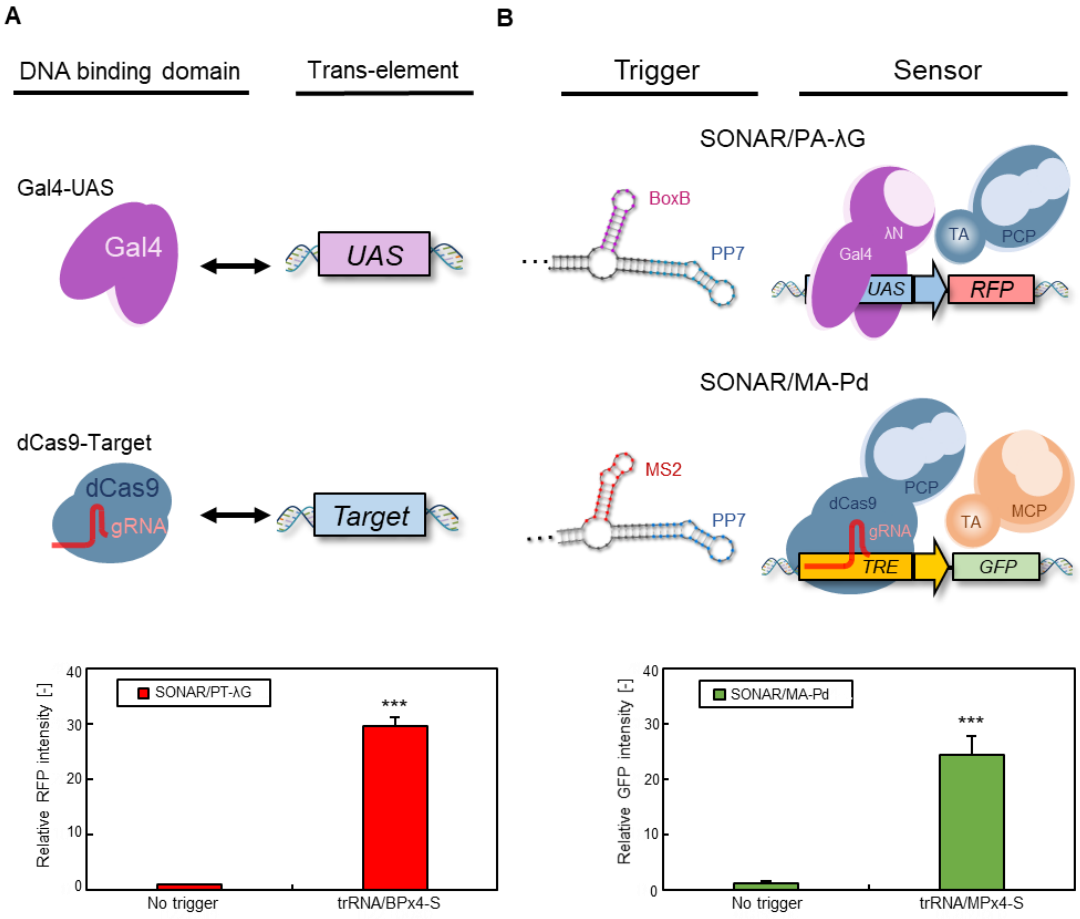


Figure 21. RNA inducible transgene expression using different DNA binding domain. (A) Schematic illustration of DNA binding domain and its binding element. Gal4 can bind to UAS element when dCas9 can bind to any designed target with the guide of gRNA. (B) SONAR/PA-λG, SONAR system using fusion protein of λN-Gal4 and PCP-TA. (C) Relative RFP intensity of the transfected CHO cells using SONAR/PA-λG and trRNA/BPx4-S. (D) Relative GFP intensity of the transfected CHO cells using SONAR/MA-Pd and trRNA/MPx4-S. Gal4, yeast Gal4 regulatory protein; UAS, upstream activating sequence derived from yeast; dCas9, nuclease deactivated clustered regularly interspaced short palindromic repeat (CRISPR)-associated Cas9 protein; gRNA, guide RNA for Cas9; Target, designed DNA target for Cas9 binding. Data represents the mean ± SD, n = 3. * $p < 0.05$, ** $p < 0.01$, ** $p < 0.001$ vs. no trigger.

Double RNA detection with orthogonal manner

To determine whether plural RNA could be detected at same time without crosstalk, three fusion proteins (PCP-TA, MCP-tetR and λN-Gal4) were expressed in CHO/TRE-GFP/UAS-RFP cells for detecting RNAs with MS2-PP7 or BoxB-PP7. The fluorescence areas were determined in captured images with fluorescence microscope. Green or red fluorescence was observed in cells transfected trRNA/MPx4-S or trRNA/BPx4-S, respectively. In cells transfected with both trRNA/MPx4-S and trRNA/BPx4-S, both GFP and RFP expression have been detected, although decreased fluorescence was exhibited in cells compared to that transfected with only one type of RNA aptamer. In the absence of target RNA, no fluorescence was observed (Figure 22). These results suggested that the RNA detection system using aptamers developed in this study can detect different RNAs with orthogonal manner.

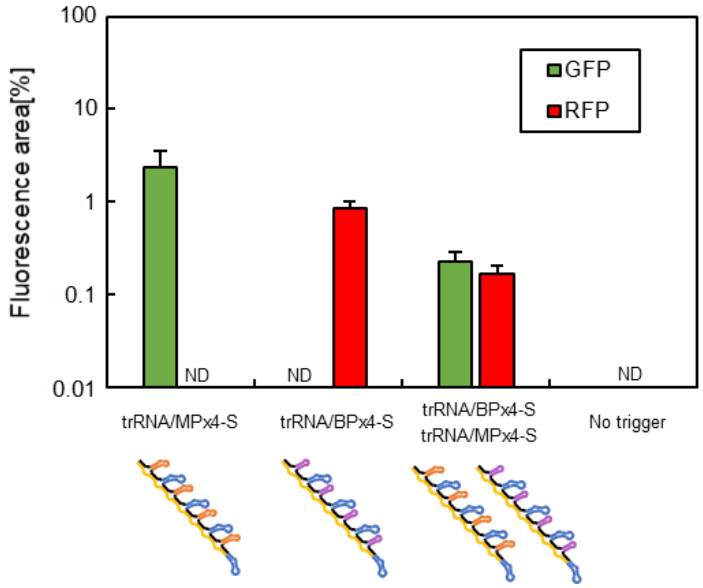


Figure 22. RNA inducible transgene expression with orthogonal manner. Fluorescence area of GFP and RFP of the transfected CHO cells using SONAR/PA-λG, SONAR/PA-MT, trRNA/MPx4-S and trRNA/BPx4-S. Data represents the mean ± SD.

3.4 Discussion

An RNA sensing module that can display the expression levels of specific RNAs with reporter gene expression was established in living cells. Expression of the reporter gene can respond to RNA tags of as little as 125 nucleotides. Thus, the RNA detection system has the potential to track the expression of not only messenger RNA but also small RNA by knocking-in aptamer tag into the target endogenous gene locus.

RNA construction with muti-valency of MS2-PP7 stem-loops enhanced reporter gene expression. A hypothetical distance/quantity model was established based on the observation that only one PCP-tetR protein recruited by PP7 is required for binding to tetO in the TRE region, and the activation of the promoter depends on both the distance between TA and tetO and the number of TA recruited to tetO. Although the physical distance and number of transcriptional activators cannot play a decisive role in microstructures with Brownian motion, the distance and number of recruitment possibly influence the probability of transcriptional activators binding to promoters. That is, the closer the bridging distance is, the greater the probability that the transcription activator will fit with the promoter DNA sequence under Brownian motion, and the more the trans-activator recruited, the greater the probability that the transcription activator will fit with the promoter DNA sequence. A mathematical model was established based on this assumption. Regression analysis proves the possibility of this model (k=4.3, R^2=0.98). Therefore, multivalent binding of trans-activator fusion proteins facilitated the RNA detection.

$$efficiency = K\sum_{i=1}^{N}(\sum_{j=1}^{N} {}^{1}/_{distance(MS2_i \rightarrow PP7_j)})$$

It is somewhat surprising that lower leak expression of GFP was observed in cells expression PCP-tetR and MCP-TA when compared with untransfected cells. This intriguing finding might be a result of sterically interfering the binding of RNA polymerase or auxiliary transcription factors by tetR **[86]**.

For detection of specific RNA via gene expression regulation, Sirius scaffold

benefit the transcriptional activation and 200-fold gene activation was achieved using MS2-PP7 aptamer array with Sirius scaffold. The on/off ratio is higher than that induced by the Cas12a gRNA processing using toehold-mediated strand displacement that reported previously (42-fold gene activation in HEK293 cells and ~100-fold gene repression in *E. coli*) **[63]**. Different RNA aptamer constructs exhibit gradient levels of gene activation. Therefore, sensitivity of this RNA sensor is adjustable by tuning aptamer array structure according to the level of detected RNA for suiting different application conditions. It can also be applied as a tunable resistor in gene circuits.

This study found that staggered type aptamer array exhibits the highest on/off ratio and significantly lower transcriptional activation was observed in trRNA/MPx2-S when compared with trigger RNAs with four MS2-PP7. These results are consistent with our hypothesized distance/quantity model ($k=24.3$, $R^2=0.98$). In trRNA/MPx4-S, any selected PP7 has a 75% probability of being adjacent to two MS2s and a 25% probability of that to one MS2. In contrast, any PP7 in trRNA/MMPPx2-S or trRNA/Mx4-Px4-S cannot be adjacent to two MS2s.

With mRNA detection strategy, mRNA expression controlled by different promoters induced different expression level of GFP using SONAR. Therefore, SONAR has the potential to be applied for detecting mRNA level and monitoring the activity of endogenous promoters.

A serious of RNA-binding domains was characterized for RNA responsive gene expression. RNA binding domains including MCP, PCP, COM and λN. Therefore, bacteriophage coat proteins can be replaced by others without affecting the detecting ability of target RNA. Recently, it has been reported that CRISPR-associated protein 13 (Cas13) can bind to arbitrary RNA sequence and have been applied for live-cell RNA imaging has been reported. Since the RNA-binding domain of SONAR can be replaced, Cas13 is a potentially RNA-binding module for SONAR RNA sensor construction which can be applied for arbitrary RNA detection.

Com arrays with double strands linkers using Sirius that can efficiently recruit proteins were constructed, which have not been achieved in previous studies **[87]**, and provide new solutions for RNA imaging other than MS2, PP7, λN22, U1A and BglG.

Nature RNA sequence was detected using SONAR system with toehold switch. Although significant transcription activation was observed with trigger, the fold-change was still not high when compared with that using toehold switch gRNA with dCas9 trans-activator shown in previous study due to the significant level of the background leak**[88]**. It has been proved that optimizing toehold switch using shorter toehold, shorter target RNA sensitive domain or introduce some mismatch between RNA sensitive domain and pp7 stem region may facilitate the on/off ration of arbitrary RNA detection**[63, 88]**. Given the high dynamic range observed in SONAR sensor, SONAR-derived arbitrary RNA sensors have greater potential to achieve efficient endogenous RNA detection after optimization.

To extend the application of SONAR, variety of DNA binding domain was characterized. DNA binding domain including dCas9, tetR and Gal4 could initiate RNA responsive gene expression using SONAR. Modulating endogenous gene expression can manipulate cell differentiation for tissue engineering or cellular metabolism for cell production optimizing**[46, 89]**. Dead-Cas9-based transcription activation system has been developed as a platform for reprogramming signal pathway networks in cells and controlling cell function and fate**[75, 87]**. However, the developed gene circuits can only accept artificial signals as switches for regulating expression of endogenous genes. RNA sensor with dCas9 module can achieve inner cell signal-controlled cell fate modulating by engineering gene circuits with RNA sensor that inducing endogenous gene expression depends on other endogenous gene transcription, connecting not only the output but also input to natural cellular signaling pathways for establishing artificial gene network.

In summary, there are no previous reports on RNA responsive gene expression based on phage aptamers. a novel RNA detection platform was designed for engineering gene circuit using RNA dependent trans-activation complex formation strategy. With this strategy, RNA binding module, DNA binding module and transcriptional regulation module can be customized for certain application. Significant reporter gene activation was confirmed by the induction of messenger RNAs with or without aptamer array. Such aptamer-based RNA detection system may serve as new parts for gene circuit

construction.

Chapter 4

Development of an inducible gene expression system using an artificial transactivator

4.1 Introduction

Chinese hamster ovary (CHO) cells are extensively used for producing biopharmaceuticals such as therapeutic antibodies. Reduced culture temperature may provide a benefit to improve the productivity and nutrient availability, indicating the shift of energy consumption from cell growth toward product synthesis. Here, a cold-inducible gene expression system was developed based on the tetracycline-regulated artificial trans-activator system for recombinant antibody production.

CHO cells are an epithelial cell line which is derived from ovary of Chinese hamster **[90, 91]**. As a mammalian cell, CHO cells can produce biopharmaceuticals with appropriate post-translational modification in a very high efficiency. However, the productivity of biopharmaceuticals in CHO cells is still not enough when compared with the great demand of biopharmaceuticals drugs and the high cost of culturing CHO cells. Thus, there is a great demand of improving the production process, reducing costs and increasing production for biopharmaceutical production.

Varity of synthetic components have been applied for promoting the production of biopharmaceutical in CHO cells. An accumulative gene integration system was developed for targeted integration and amplification of biopharmaceutical gene using Cre/loxP system **[26, 54, 55, 92, 93]**. An engineered CHO cell line was established by target introducing scFv-Fc antibody gene to hprt locus using clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 and CRISPR-mediated precise integration into target chromosome (CRIS-PITCh) systems. Regulating gene expression using synthetic components may benefit the production of biopharmaceuticals **[33]**. Therefore, the purpose of this experiment is to develop a novel gene element for regulating gene expression using trans-activator.

To regulate the expression of transgene, nature gene element was optimized and assembled for generating engineered elements. Two types of elements are found in nature cells affecting the gene transcription, cis-acting elements and trans-acting elements. The trans-acting elements encode proteins consists of DNA-binding domain and transcriptional regulation domain. Trans-acting elements could bind to cis- acting element and regulating nearby gene activity. Tetracycline inducible transcriptional activator (*tTA*) including DNA binding domain, tetR, and transcriptional activation domain, VP64, is a kind of trans-acting element, which could bind to TRE promoter and activate the expression **[30, 43, 86]**. However, as described in Chapter 3, transcriptional activation could also be induced by other trans-activators. Identified trans-activator domains in mammalian cells was shown in Table 7.

Table 7 Identified trans-activator domain and fusion protein in mammalian cells

Activation domains	Full name
VP64	VP64 transactivator
p65	p65 subunit of NF-κB
Rta	Epstein-Barr virus R transactivator
AD2	apolipoprotein E
CR3	integrin alpha M
GATA4	GATA binding protein 4
p53	tumor protein p53
SP1	trans-acting transcription factor 1
MYOD	myogenic differentiation protein
MEF2C	myocyte enhancer factor 2C
TAX	tax protein
PPARγ	peroxisome proliferator activated receptor gamma
Gal4	galactose-responsive transcription factor GAL4
VP192	VP192 transactivator
HSF1	heat shock transcription factor 1
P300	E1A binding protein p300
VPR	Fusion of VP64-p65-Rta

Nat Methods 12(4):326-328 2015

Drug inducible expression regulation systems including tetracycline expression regulatory system have developed using trans-activators. However, drug inducible expression regulation system can be hardly applied for industrial production of recombinant protein due to the cost of drug and purification. Therefore, an additive-free gene expression induction system was required for enhancing the biopharmaceuticals production.

Regulating gene expression by controlling culture temperature may benefit the biopharmaceuticals production. We have developed a heat inducible gene expression

regulation system for gene therapy **[5, 34, 37, 94]** and a hypoxia inducible translation controlling system for tissue engineering **[39]**. In contrast, low temperature can be used not only to regulate cell expression, but also to regulate cell proliferation to control cell nutrient distribution.

Therefore. the purposed of this experiment is developing a cold-inducible gene expression system by using artificial genetic circuit based on tetracycline-regulated transactivator system for recombinant antibody production (Figure 23).

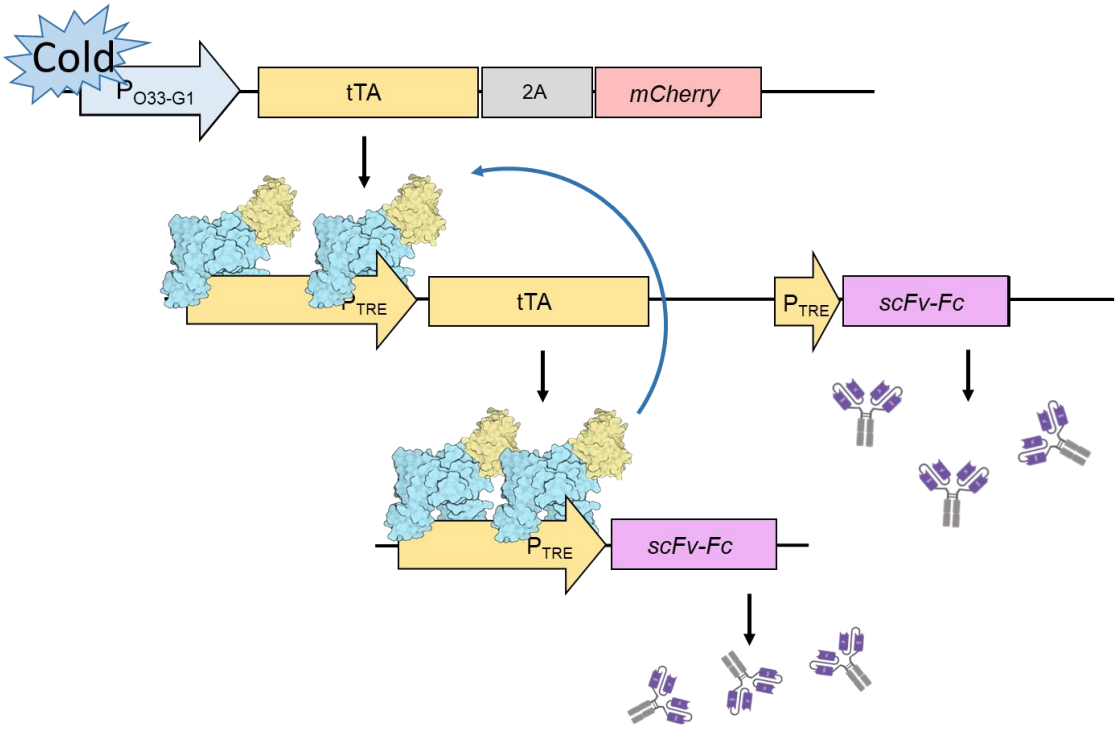


Figure 23. Schematic representation of temperature shift-inducible scFv-Fc expression system.

4.2 Materials and methods

Cell culture

CHO-K1 cells (RIKEN, Tsukuba, Japan) were maintained in same condition described in Chapter 2.

Plasmid construction

To sense temperature shift, tetracycline inducible transcriptional activator (*tTA*) and red fluorescent protein (mCherry) co-expression unit was chemically synthesized. O33G1 gene promoter was predicted by genomic database (http://cgr-referencegenome.boku.ac.at/). Specific primers were designed upstream and downstream of the predicted promoter region, ligated with SalI and EcoRI tags, and specifically amplified the promoter using KOD-plus-neo. Tetracycline-regulated transactivator (*tTA*) gene co-expressed with *mCherry* (*tTA-2A-mCherry*) under the control of cold inducible promoter (*O33-G1*) was introduced into piggyBac-based transposon vector (pPB) harboring zeocin resistant gene to generate pPB/*tTA*-Red (Figure 24).

Figure 24. Plasmid design for cold sensor.

DNA fragments encoding *tTA* and copGFP-T2A-Puror expression units under the control of TRE promoters were chemically synthesized, respectively (Genewiz). A DNA fragment encoding TRE-*tTA* was prepared by digestion with SpeI and XhoI from pUC57/TRE-*tTA*2s. The fragment was ligated with another DNA fragment encoding TRE-GFP-Puro by digestion with XhoI and BstXI from pUC57/TRE-GFP-Puro into PiggyBac-based transposon vector (#PB513B-1, System Biosciences) by digesting with SpeI and BstXI to generate PB/[TRE-*tTA*][TRE-GFP-Puro] (PB/AT-GP).

Genomic engineered cell line generation

CHO-K1 cells (5×10^4) were harvested from 100mm dish, Resuspended with Gibco™ Opti-MEM™ (Thermo Fisher Scientific) and transfected with totally 5000ng of plasmid vector from the NEPA21 Super Electroporator (Nepa Gene Co., Ltd, Chiba, Japan) in accordance with the manufacturer's protocol. The electroporation conditions for transfection were two poring pulses at 125 V for 10 ms with 50 ms pulse interval and 10% decay rate and 5 transfer pulses at 20 V for 50 ms with 50 ms pulse interval, 50% decay rate and polarity switching. The cells were seeded in 6-well plates with 2 mL of F-12 culture medium. For transient expression, the cells were seeded in three wells for each transfection condition and observed by fluorescence microscopy and cell sorter (Sony Biotechnology Inc, Tokyo, Japan) 48 hours after transfection.

ScFv-Fc productivity

Specific scFv-Fc productivity was measured as described previously in Chapter 2.

4.3 Results

Cold-inducible gene expression

To regulate the production of biopharmaceuticals, cold responsive inducible gene expression system was developed by discovering nature cold inducible genes and synthesize expression unit containing cold inducible promoter and transgene (Figure 25).

To discover nature cold inducible element in CHO cells, transcriptome of CHO-K1 cells cultured at 33°C and 37°C were analyzed by microarrays for screening cold inducible genes. According to the result of DNA microarrays, several genes were enhanced by low temperature. The expression level of three alternative genes, *O33-G1, O33-G2* and *O33-G3*, was confirmed by reverse-transcription PCR and real-time quantitative reverse-transcription PCR. Both alternative genes were enhanced by low temperature. Especially the expression of *O33-G1* gene was observed to be enhanced over 500-fold in 33°C. Apoptosis of CHO-K1 cells was not observed when cultured in 33°C, which may benefit the protein production (Figure 26).

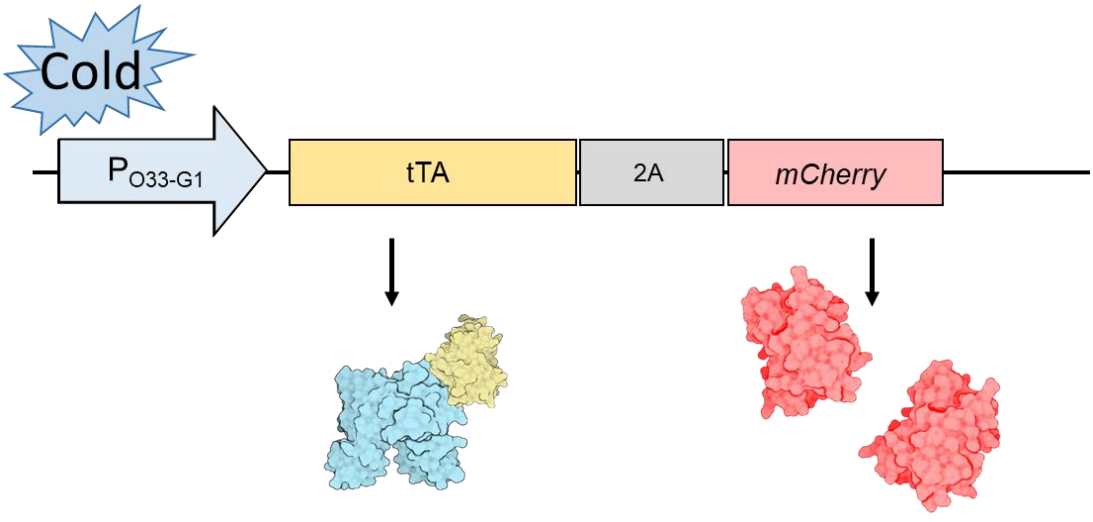


Figure 25. Schematic representation of cold sensor; (A). Discovery of cold inducible gene element; (B) Integration of temperature shift inducible gene transcription activation system.

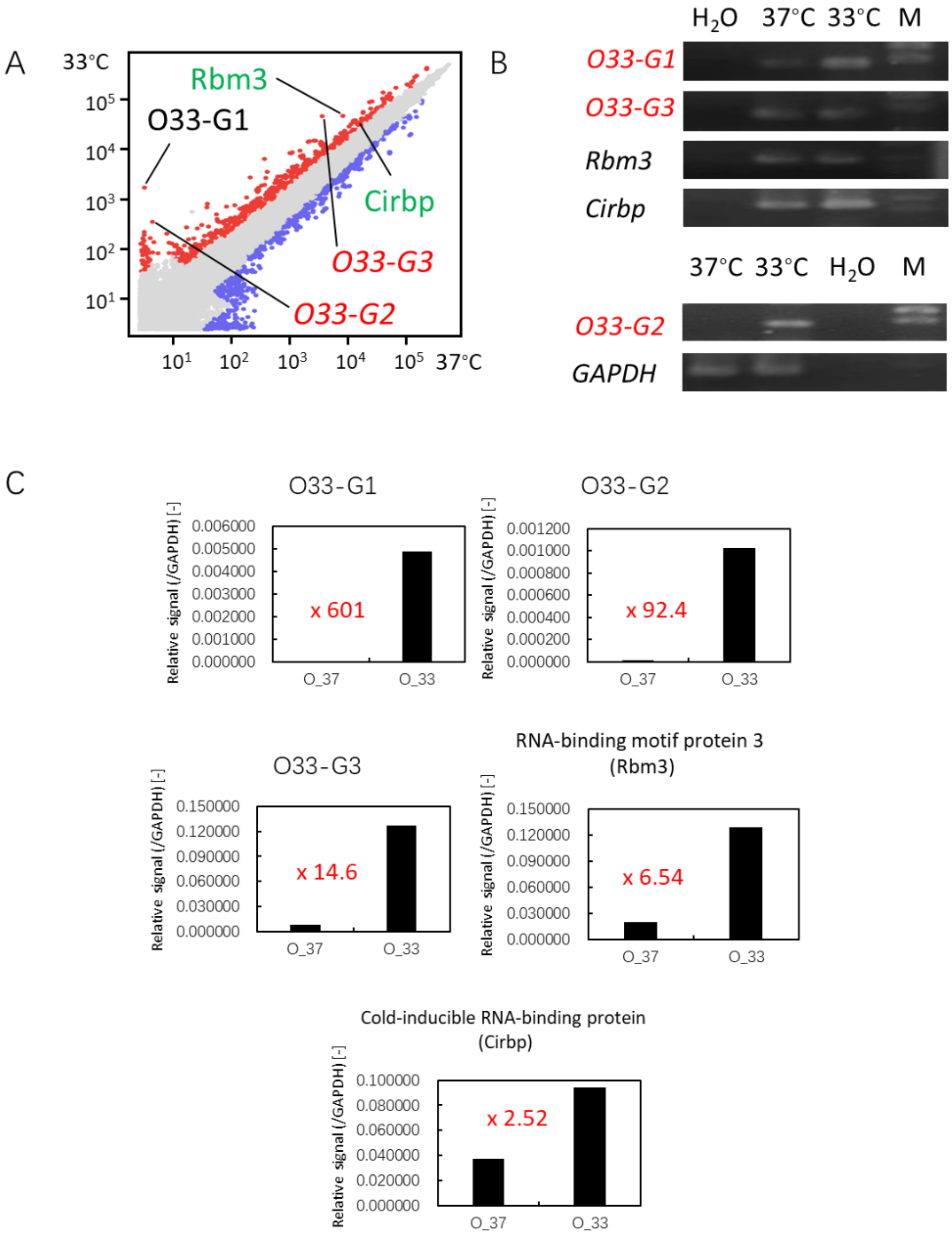


Figure 26. Discovery of cold inducible gene element (A) Scatterplot of micro array data; (B) Transcription level of cold inducible gene using reverse transcription PCR; (C)Histogram of micro array data.

To evaluate the cold inducible promoter, a temperature sensor for transferring temperature signal to transcriptional activation signal was developed. *O33-G1* promoter was chosen as the cold inducible promoter. CHO cells were integrated with *tTA*-2A-mCherry gene under the control of with *O33-G1* promoter for establishing cold responsive recombinant cell line. By cell sorting of temperature response in terms of red fluorescence, the clone showing higher inducibility in 33°C was screened. Higher expression of mCherry at 33°C were observed in developed clones compared with that at 37°C using cell sorting analysis. Established CHO cell line exhibited a significant enhancement of mCherry gene expression in 33°C (Figure 27). This result indicated that the developed cold sensor induced a temperature gene expression activation.

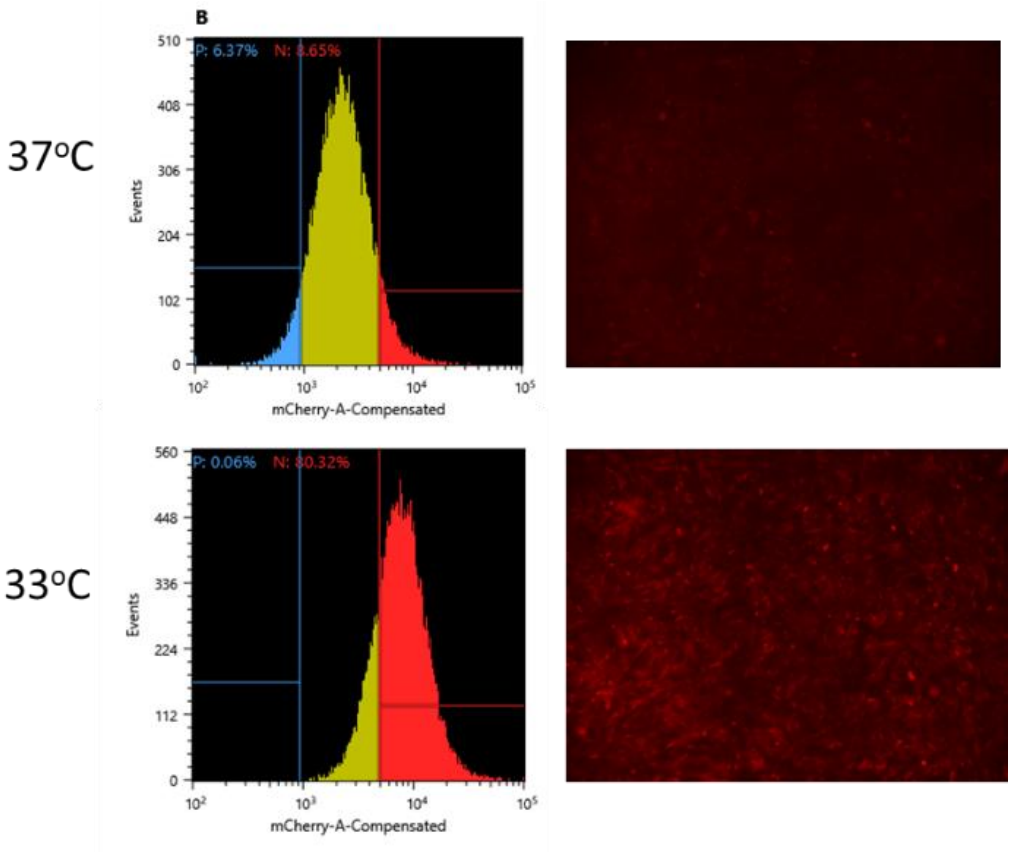


Figure 27. Cold inducible expression of mCherry gene in CHO/*O33-G1* cells.
Left: mCherry flowcytometry of CHO/*O33-G1* cells cultured in 33°C or 37°C.
Right: fluorescence images of CHO/*O33-G1* cells cultured in 33°C or 37°C.

Transactivator-mediated scFv-Fc production

To establish biopharmaceutical producing cell line, inducible expression unit of scFv-Fc was designed. The cells integrating *tTA*-2A-mCherry gene under the control of with *O33-G1* promoter exhibited higher expression of mCherry at 33°C culture condition compared with that at 37°C. Therefore, a cold responsive CHO cell lines consists of cold inducible trans-activator expression unit have been established. Expression unit for controlling scFv-Fc expression were designed by synthesizing TRE controlled scFv-Fc expression unit, which was integrated into CHO cells containing cold inducible *tTA* expression unit for establish scFv-Fc inducible expressing cell line (CHO/CSscFv) (Figure 28).

When the CHO/CSscFv cells were cultured at 37°C and 33°C culture condition, the scFv-Fc productivity at 33°C (24 pg/(cell•day)) was 3-fold higher compared with that at 37°C (8 pg/(cell•day)) (Figure 29) after 5 days culture. CHO cell lines producing recombinant antibody production was established using a synthetic cold-inducible gene expression system which can be utilized for the production of biopharmaceutical proteins. This result indicates that gene modules can be simply combined with each other for achieving complex function.

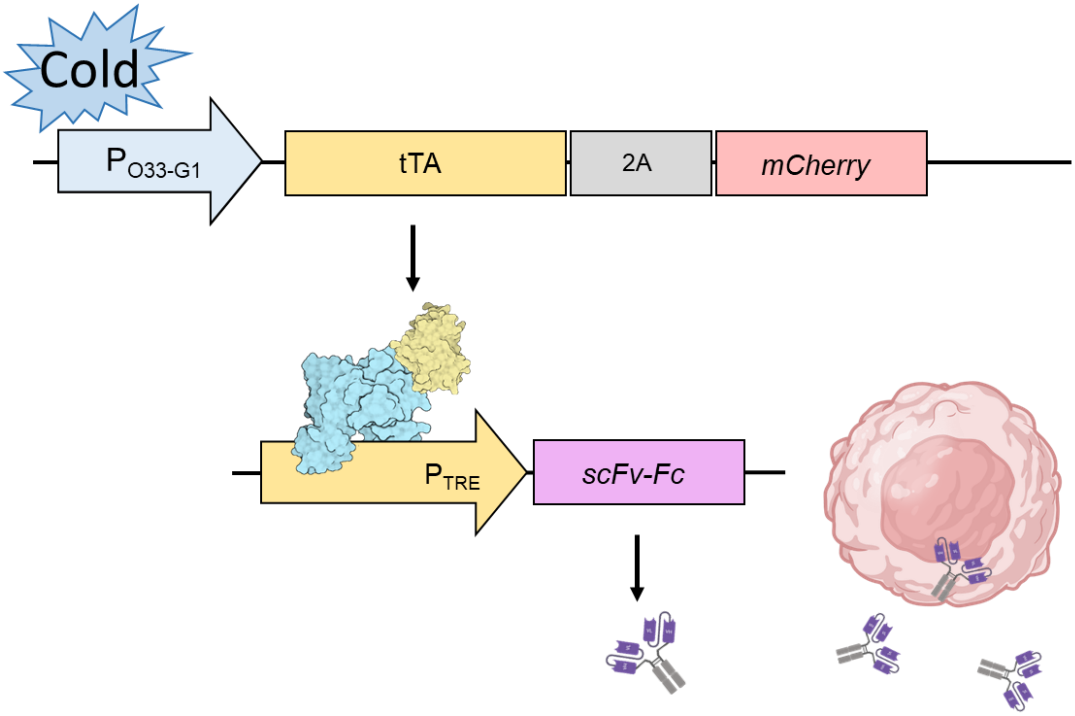


Figure 28 Schematic representation of trans-activator-mediated scFv-Fc production.

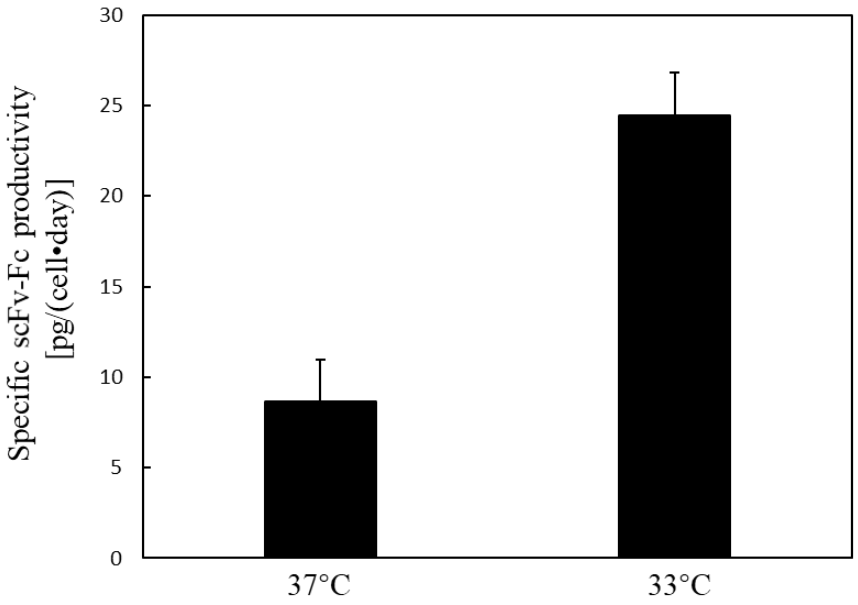


Figure 29. Specific scFv-Fc productivity of temperature shift inducible scFv-Fc expression system.

Positive-feed-back mediated expression for transcriptional enhancer

In order to amplify the productivity of scFv-Fc, a positive feedback gene circuit based on the tetracycline expression regulation system was designed. Tetracycline repressor based artificial transactivator (*tTA*) expression was controlled by a *tTA* inducible promoter (TRE) in order to generate a positive feedback loop of *tTA* expression. Positive feedback co-expression vector with GFP reporter gene were integrated into CHO cells for establishing recombinant cell line (Figure 30).

To evaluate the function of positive feedback system, the expression was activated by transfecting the *tTA* expression vector and GFP intensity was measured by flowcytometry. Recombinant CHO cells with activated positive feed-back system result in high level GFP expression (Figure 31), indicated that positive feedback expression system can effectively induce high level expression level of GFP. The expression of GFP was effectively inhibited by tetracycline analogs, doxycycline, which makes the expression regulation system based on positive feedback has more application potential.

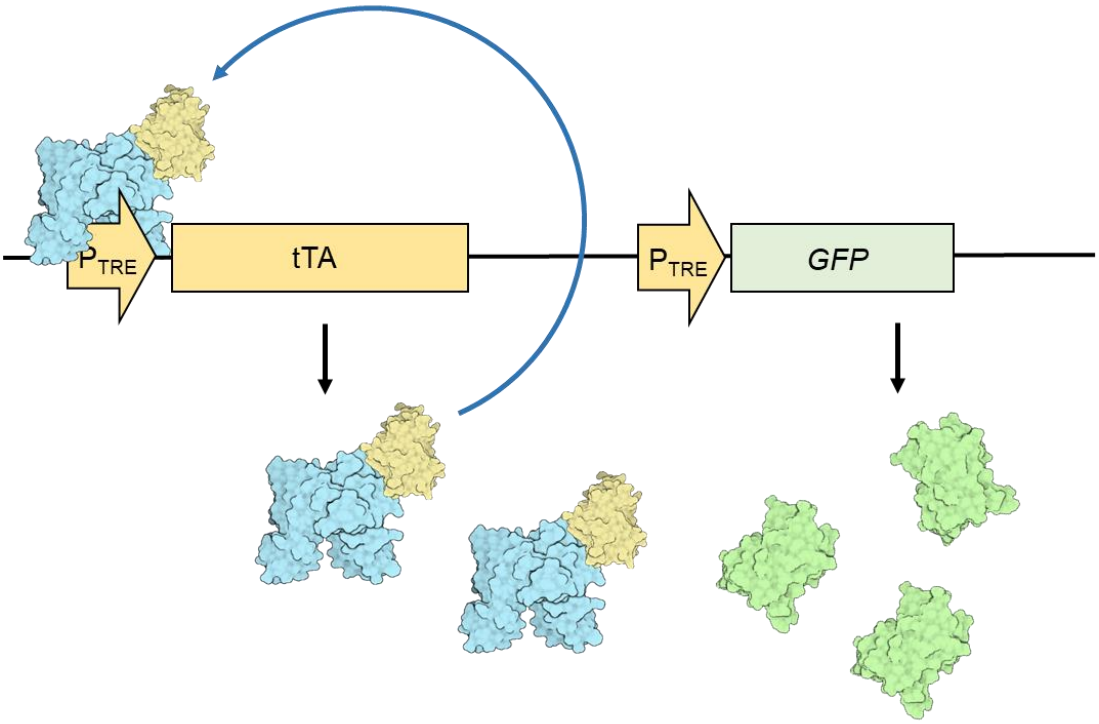


Figure 30. Schematic representation of positive-feed-back expression amplification module.

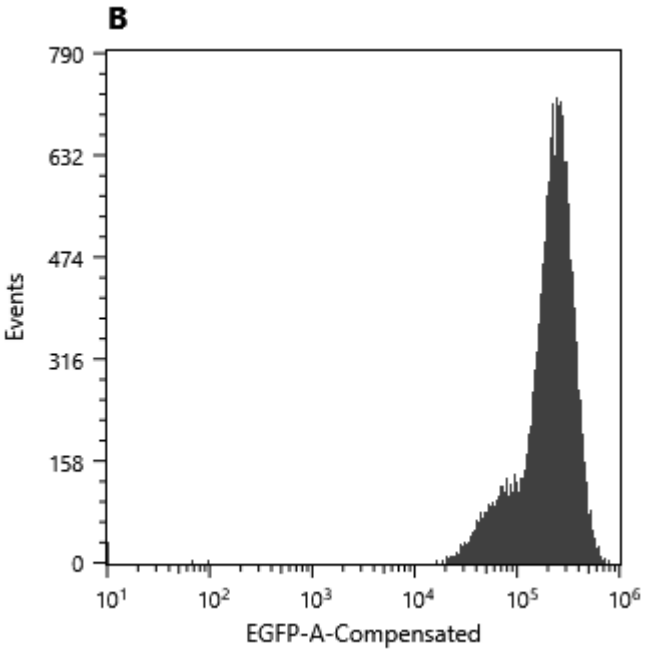


Figure 31. GFP intensity of cells introduced positive feed-back expression system.

Positive feedback expression regulation system can generate a cell memory that once activated, target gene can be expressed stably for a long time, even if the activation signal has been disappeared. Activated CHO/TRE-*tTA* recombinant cells were cultured for 28 days and GFP intensity were evaluated by cell sorter every 4 days. Although 100% of CHO/AGP#5 (dox-) cells could generate strong fluorescence after screening with puromycin for 4 days, GFP-negative cells appeared spontaneously (from 100% to 95% in 12 days). Nevertheless, most cells (over 85%) can maintain high expression of GFP in 28 days (Figure 32).

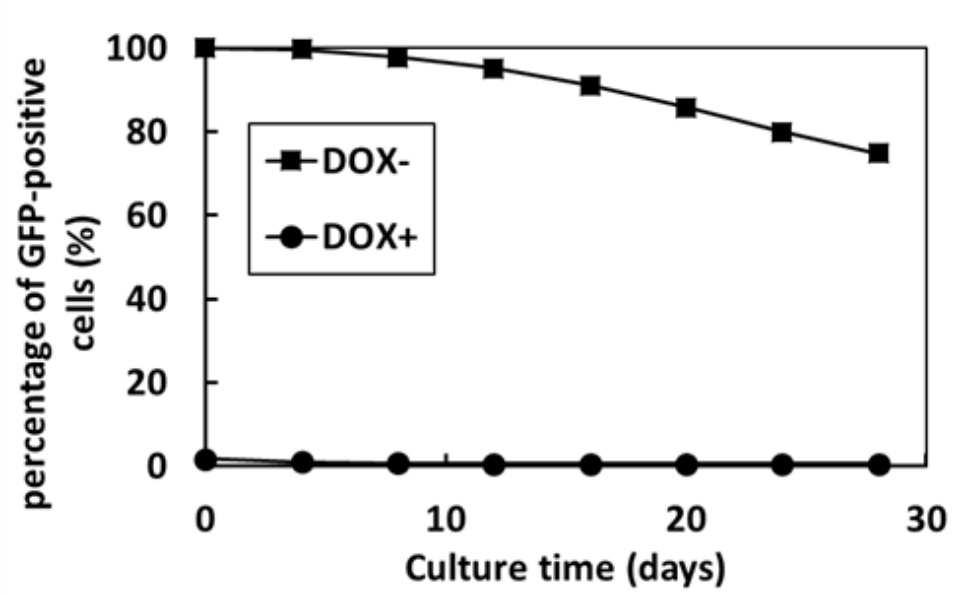


Figure 32. GFP positive cells' rate in long term culturing

Enhanced scFv-Fc production using positive-feed-back

The cold-inducible trans-activator expression unit, positive feed-back amplification system and scFv-Fc inducible expression unit have been developed for cold-inducible transgene expression, enhancement of transgene expression and biopharmaceutical production, respectively.

In order to achieve stable and efficient expression of antibodies, recombinant cells were engineered with cold inducible trans-activator expression unit, positive feed-back amplification system and scFv-Fc inducible expression unit (Figure 33).

Cell lines with positive feed-back amplification system was established (CHO/AGP#5). Temperature shift sensor containing *O33-G1* promoter driven *tTA*-2A-mCherry expression unit and scFv-Fc expression system containing TRE driven scFv-Fc expression unit was integrated into CHO/AGP #5 cells using PiggyBac transposon. The genome modified cells integrated with cold sensor, positive feed-back amplification system and scFv-Fc expression system are supposed to stably express scFv-Fc in a high level.

More than 60 PCD of scFv-Fc antibody production was observed in the recombinant cells cultured in 33 degrees (Figure 33). This result indicating that the productivity of antibodies has been significantly improved after introducing the positive feedback processing system.

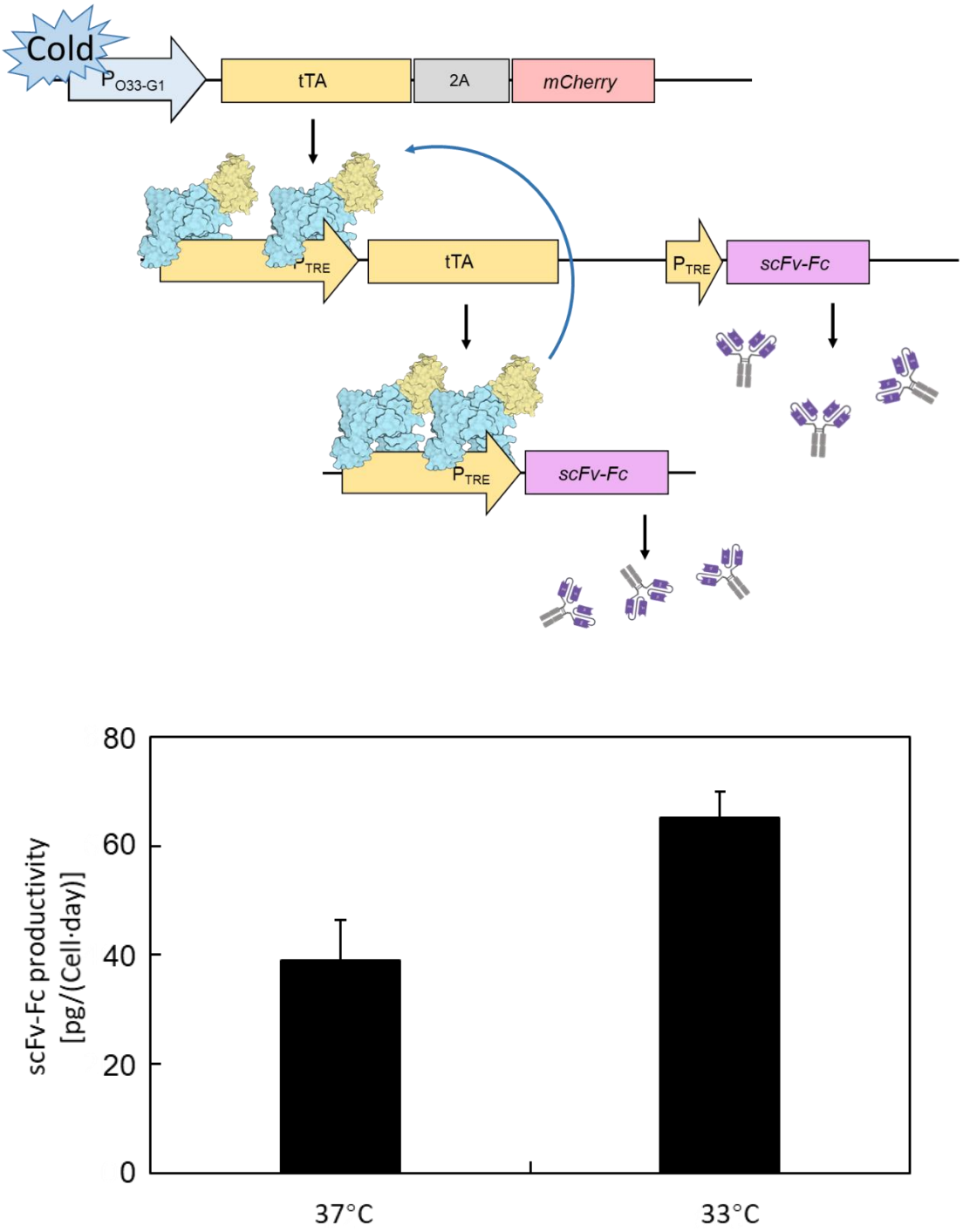


Figure 33. Cold inducible scFv-Fc expression system with PFB amplification. Schematic representation of cold inducible scFv-Fc expression system with PFB amplification. Specific scFv-Fc productivity of temperature shift inducible scFv-Fc expression system.

4.4 Discussion

A temperature shift responsive biopharmaceutical production system was developed using cold-inducible trans-activator expression unit, positive feed-back amplification system and scFv-Fc inducible expression unit.

A cold inducible promoter was discovered and assembled for engineering cold-inducible trans-activator expression unit. The *O33-G1* promoter were characterized and showed up to 500-fold activity increase when cultured in 33°C. The fold change value is 50-fold higher when compared with Ind2-Sbsn promoter reported previously, which showed a 11-fold enhancement of activity when cultured in 33°C in CHO-K1 cells **[33]**. Therefore, the cold-inducible trans-activator expression system showed a great potential for gene expression regulation.

A positive feed-back amplification system was developed for expression amplification and stabilization. Interestedly, although the transformed cells lose their GFP expression completely, the fluorescence level of GFP-positive cells didn't decrease. Only two group of cells were observed after long term culturing, GFP negative cells and GFP high expression cells. No GPF low expression cells were observed, indicating that positive feed-back system worked according to all-or-none law. This result demonstrated that once transcriptional activation energy exceeds a threshold, gene expression can be amplified to saturation using positive feedback. In long-term culture, high-speed proliferating cells produced a larger proportion of silent cells than slow-proliferating cells, suggesting that the silencing of the positive feedback system is related to cell proliferation. Cells with condensed chromosomes during division may not be able to express trans-activators, resulting in the number of transcriptional activators in some cells being less than the threshold, resulting in gene silencing. Slowing cell proliferation can delay the emergence of silent cells and low temperature culture can reduce the proliferation rate of cells. Therefore, the positive feedback gene expression system applied for cold culture has the potential for long-term expression of transgenes.

The temperature-shift responsive biopharmaceutical production system induced

more than 60 PCD of scFv-Fc antibody production in recombinant CHO cells. This system can effectively regulate the production of biopharmaceuticals and achieve a balance between cell proliferation and drug production. Notably, the production of scFv-Fc in CHO cells using positive feedback amplification system was comparable to our previous study using the AGIS system, without the requiring of complicated introduction steps **[55]**.

In summary, a novel inducible gene expression system was developed for biopharmaceutical production. A cold-inducible trans-activator expression unit was developed for cold-inducible transgene expression. A positive feed-back amplification system was constructed in CHO cells for enhancement of transgene expression. A scFv-Fc inducible expression unit have been applied for inducible biopharmaceutical production. High expression of scFv-Fc was observed in recombinant CHO cells.

Chapter 5

Summary

Engineered animal cells are cells that have been genetically modified or manipulated to perform specific functions or express specific proteins. Genetically engineered animal cells are used in a wide range of applications including biotechnology, medicine, and research.

In this thesis, synthetic biological approaches were used to develop three techniques for establishing engineered animal cells. To modify animal cells engineering strategy, artificial gene delivery systems were designed to integrate DNA fragments containing artificial gene expression system into the genome. And the artificial gene expression systems initiate and regulate the expression of recombinant proteins to perform specific functions. Natural genetic modules derived from different species were modified and assembled to generate artificial biological systems.

LINE-1 gene delivery system was developed to integrate muti-copies of gene expression systems into the genome. Artificial gene expression systems including RNA-responsive gene expression regulation systems and cold-inducible gene expression systems have been developed to control the functions of engineered animal cells.

The retrotransposon-based gene delivery system successfully delivered the scFv-Fc gene expression unit. The presence of the scFv-Fc gene was confirmed in the genome of the engineered cell line, and the production of scFv-Fc protein was detected in the culture medium. These results indicate that the LINE-1-based gene delivery system can be effectively utilized for biopharmaceutical production by generating biopharmaceutical-producing cells (Figure 34).

The gene expression regulation systems for RNA detection have potential applications in gene therapy. In the future, it could be used to detect cancer- or virus-specific RNA, and activate the expression of suicide or antibody genes to eliminate cancer cells or viruses. However, in this study, the RNA detection system was designed

to detect specific synthetic RNAs containing phage-derived stem loops. To apply this system to gene therapy, it is essential to detect natural RNAs. Therefore, an RNA strand displacement strategy was designed to allow detection of any natural RNA sequence by designing an appropriate probe RNA. With the natural RNA detection system, the gene expression regulation system can be applied for gene therapy (Figure 35).

The cold-inducible gene expression system is another type of gene expression regulation system, which consists of a cold-inducible transactivator expression unit, a positive-feedback expression amplification unit and an scFv-Fc expression unit. The transcription of the scFv-Fc gene is initiated at 33°C and high levels of scFv-Fc protein production were detected in the culture medium. Therefore, the cold-inducible gene expression system can be applied for cost-effective production of biopharmaceuticals by controlling the cell state.

In summary, three artificial biological systems have been developed to create engineered animal cells using synthetic biological techniques. These engineered animal cells may be used for gene therapy and biopharmaceutical production.

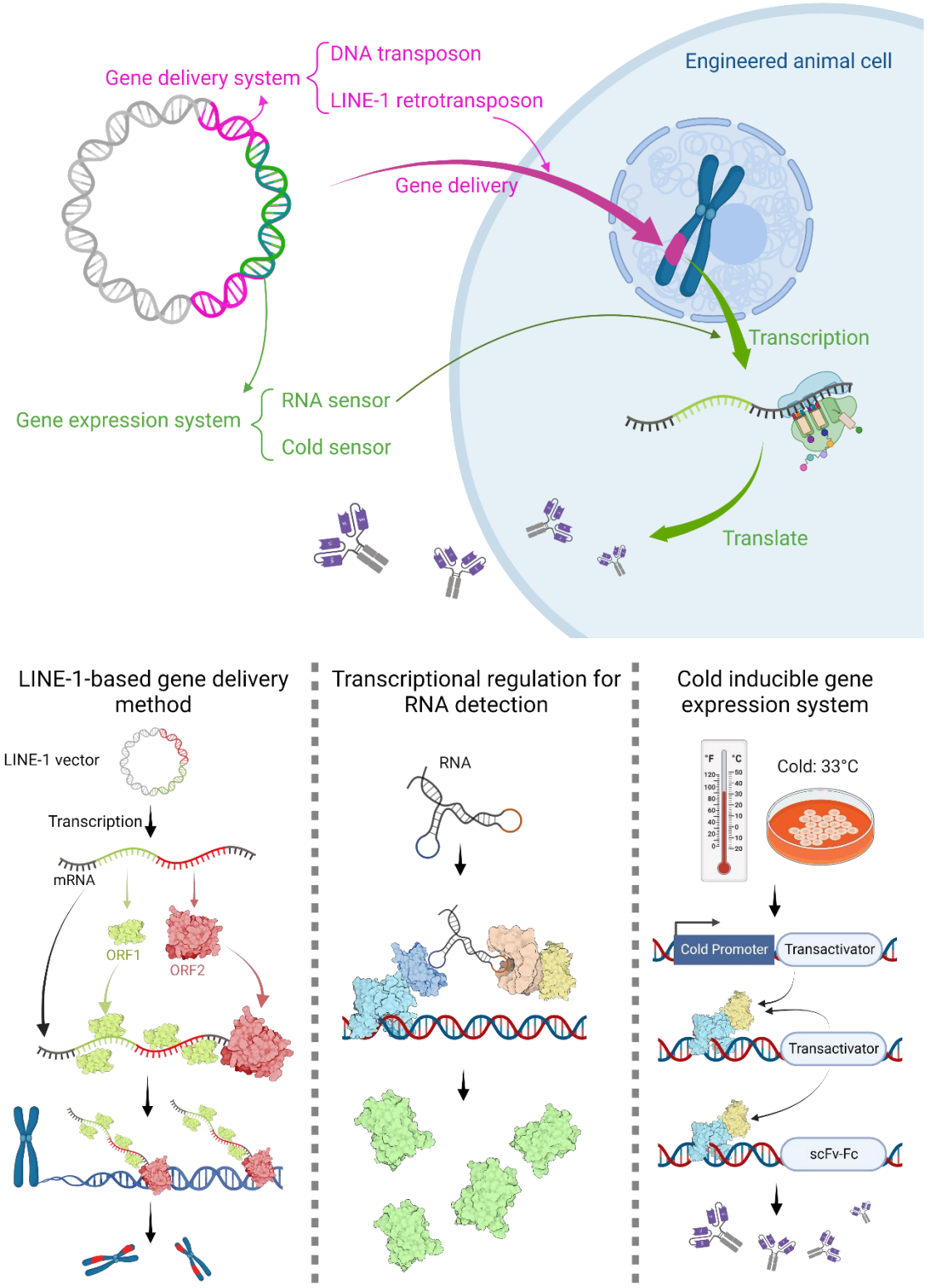


Figure 34. Schematic representation of animal cell engineering technology.

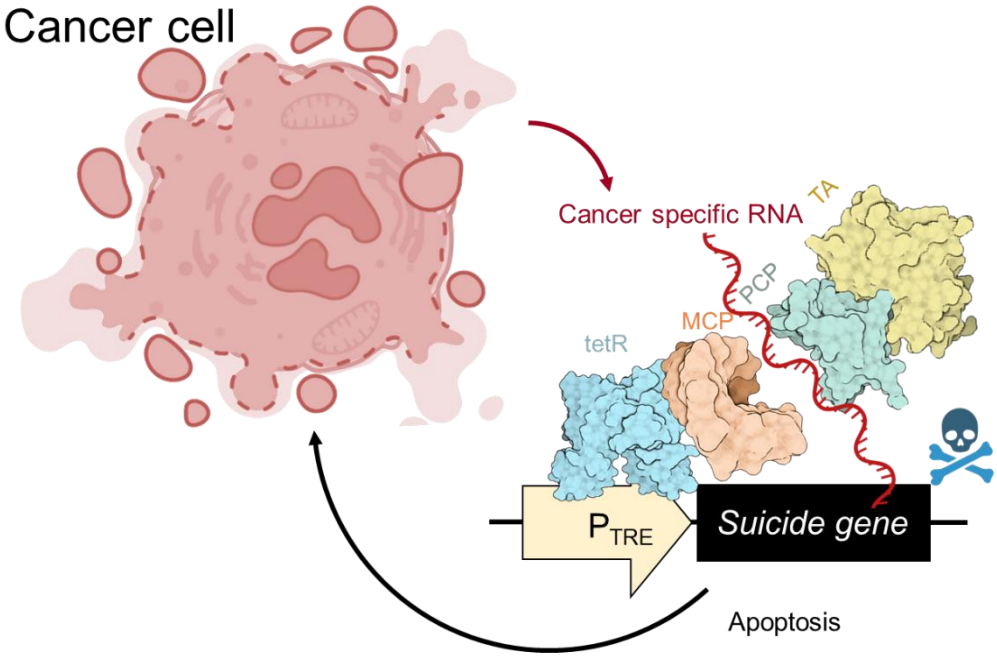


Figure 35. Schematic representation of RNA sensor-based gene therapy.

Acknowledgements

First and foremost, I would like to show my deepest appreciation to my advisor Prof. Kamihira Masamichi for his support and encouragement. He imparted both knowledge and inspiration to me, and also demonstrated how to cultivate independence as a researcher. With the lessons learned from him, I aim to continue my growth and pursue excellence as a researcher, just as he did.

I would like to further extend my thanks to Assoc. prof. Yoshinori Kawabe who provided guidance and suggestion during my experiment.

I extend my gratitude to Prof. Ijima Hiroyuki and Prof. Katakura Yoshinori for their invaluable support in reviewing my paper and offering insightful suggestions and comments, which greatly contributed to the improvement of this thesis.

I also wish to give my thanks to all the members of Lab 3 including Ms. Murakami Mai, Mr. Kuno and Dr. Yoshioka Kantaro. They help me a lot in my research and life in Japan.

Figures in this thesis was created with BioRender.com and publication rights were reserved.

Protein 3D structure data was downloaded from Protein Data Bank.

Finally, I would like to thank my parents and my best friend Ji. They give me support and encouragements through my life.

References

[1] Verbic, A., et al. (2021) *A guide to the design of synthetic gene networks in mammalian cells*. FEBS J ***288***, 5265-5288.

[2] Abudayyeh, O. O., et al. (2017) *RNA targeting with CRISPR-Cas13*. Nature ***550***, 280-284.

[3] Jusiak, B., et al. (2016) *Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology*. Trends Biotechnol ***34***, 535-547.

[4] Huang, H., et al. (2019) *Oncolytic adenovirus programmed by synthetic gene circuit for cancer immunotherapy*. Nat Commun ***10***, 4801.

[5] Kitano, H., et al. (2022) *HepG2-Based Designer Cells with Heat-Inducible Enhanced Liver Functions*. Cells ***11***.

[6] Xie, Z., et al. (2011) *Multi-input RNAi-based logic circuit for identification of specific cancer cells*. Science ***333***, 1307-1311.

[7] Auslander, S., et al. (2012) *Programmable single-cell mammalian biocomputers*. Nature ***487***, 123-127.

[8] Elowitz, M. B., and Leibler, S. (2000) *A synthetic oscillatory network of transcriptional regulators*. Nature ***403***, 335-338.

[9] Moon, T. S., et al. (2012) *Genetic programs constructed from layered logic gates in single cells*. Nature ***491***, 249-253.

[10] Hayashida, G., et al. (2019) *Differential genomic destabilisation in human cells with pathogenic MSH2 mutations introduced by genome editing*. Exp Cell Res ***377***, 24-35.

[11] Ono, A., et al. (2015) *DNA damage-responsive transgene expression mediated by the p53 promoter with transcriptional amplification*. J Biosci Bioeng ***120***, 463-466.

[12] Watson, J. D., and Crick, F. H. (1974) *Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in Nature, number 4356 April 25, 1953*. Nature ***248***, 765.

[13] Hafner, A., et al. (2020) *Quantifying the Central Dogma in the p53 Pathway in Live Single Cells*. Cell Syst ***10***, 495-505 e494.

[14] Gibson, D. G., et al. (2010) *Creation of a bacterial cell controlled by a chemically synthesized genome*. Science ***329***, 52-56.

[15] Camacho-Soto, K., et al. (2014) *Ligand-gated split-kinases*. J Am Chem Soc ***136***, 3995-4002.

[16] Meng, F., and Ellis, T. (2020) *The second decade of synthetic biology: 2010-2020*. Nat Commun ***11***, 5174.

[17] Chakrabarty, P., et al. (2022) *Combinatorial physical methods for cellular therapy: Towards the future of cellular analysis?* J Control Release

10.1016/j.jconrel.2022.12.038.

[18] Zhao, S., et al. (2016) *PiggyBac transposon vectors: the tools of the human gene encoding*. Transl Lung Cancer Res ***5***, 120-125.

[19] Suematsu, M., et al. (2022) *PiggyBac Transposon-Mediated CD19 Chimeric Antigen Receptor-T Cells Derived From CD45RA-Positive Peripheral Blood Mononuclear Cells Possess Potent and Sustained Antileukemic Function*. Front Immunol ***13***, 770132.

[20] Wu, W., et al. (2017) *AAV-CRISPR/Cas9-Mediated Depletion of VEGFR2 Blocks Angiogenesis In Vitro*. Invest Ophthalmol Vis Sci ***58***, 6082-6090.

[21] Maddalo, D., et al. (2015) *Corrigendum: In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system*. Nature ***524***, 502.

[22] Maddalo, D., et al. (2014) *In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system*. Nature ***516***, 423-427.

[23] Arsenijevic, Y., et al. (2022) *Lentiviral Vectors for Ocular Gene Therapy*. Pharmaceutics ***14***.

[24] Cavalieri, V., et al. (2018) *Non-Primate Lentiviral Vectors and Their Applications in Gene Therapy for Ocular Disorders*. Viruses ***10***.

[25] Sakuma, T., et al. (2012) *Lentiviral vectors: basic to translational*. Biochem J ***443***, 603-618.

[26] Kawabe, Y., et al. (2012) *Repeated integration of antibody genes into a pre-selected chromosomal locus of CHO cells using an accumulative site-specific gene integration system*. Cytotechnology ***64***, 267-279.

[27] Wang, X., et al. (2018) *Cre-Mediated Transgene Integration in Chinese Hamster Ovary Cells Using Minicircle DNA Vectors*. Biotechnol J ***13***, e1800063.

[28] Shi, M., et al. (2020) *Targeted knock-in into the OVA locus of chicken cells using CRISPR/Cas9 system with homology-independent targeted integration*. J Biosci Bioeng ***129***, 363-370.

[29] Zheng, F., et al. (2021) *LINE-1 vectors mediate recombinant antibody gene transfer by retrotransposition in Chinese hamster ovary cells*. Biotechnol J ***16***, e2000620.

[30] Das, A. T., et al. (2016) *Tet-On Systems For Doxycycline-inducible Gene Expression*. Curr Gene Ther ***16***, 156-167.

[31] Armbruster, B. N., et al. (2007) *Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand*. Proc Natl Acad Sci U S A ***104***, 5163-5168.

[32] Gardner, T. S., et al. (2000) *Construction of a genetic toggle switch in Escherichia coli*. Nature ***403***, 339-342.

[33] Nguyen, L. N., et al. (2020) *Bioinformatic Identification of Chinese Hamster Ovary (CHO) Cold-Shock Genes and Biological Evidence of their Cold-Inducible Promoters*. Biotechnol J ***15***, e1900359.

[34] Yamaguchi, M., et al. (2014) *Heat-inducible gene expression system by applying alternating magnetic field to magnetic nanoparticles*. ACS Synth Biol ***3***, 273-279.

[35] Kennedy, M. J., et al. (2010) *Rapid blue-light-mediated induction of protein interactions in living cells*. Nat Methods ***7***, 973-975.

[36] Ochoa-Fernandez, R., et al. (2016) *Optogenetics in Plants: Red/Far-Red Light Control of Gene Expression*. Methods Mol Biol ***1408***, 125-139.

[37] Ito, A., et al. (2019) *Magnetically triggered transgene expression in mammalian cells by localized cellular heating of magnetic nanoparticles*. J Biosci Bioeng ***128***, 355-364.

[38] Pan, Y., et al. (2018) *Mechanogenetics for the remote and noninvasive control of cancer immunotherapy*. Proc Natl Acad Sci U S A ***115***, 992-997.

[39] Masumoto, S., et al. (2021) *Hypoxia-responsive expression of vascular endothelial growth factor for induction of angiogenesis in artificial three-dimensional tissues*. J Biosci Bioeng ***132***, 399-407.

[40] Lee, D., et al. (2017) *A calcium- and light-gated switch to induce gene expression in activated neurons*. Nat Biotechnol ***35***, 858-863.

[41] Nguyen, N. T., et al. (2018) *Rewiring Calcium Signaling for Precise Transcriptional Reprogramming*. ACS Synth Biol ***7***, 814-821.

[42] Bai, P., et al. (2019) *A fully human transgene switch to regulate therapeutic protein production by cooling sensation*. Nat Med ***25***, 1266-1273.

[43] Beilstein, K., et al. (2015) *Conditional control of mammalian gene expression by tetracycline-dependent hammerhead ribozymes*. ACS Synth Biol ***4***, 526-534.

[44] Higashikuni, Y., et al. (2017) *Advancing therapeutic applications of synthetic gene circuits*. Curr Opin Biotechnol ***47***, 133-141.

[45] Green, A. A., et al. (2014) *Toehold switches: de-novo-designed regulators of gene expression*. Cell ***159***, 925-939.

[46] Gao, X., et al. (2014) *Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers*. Nucleic Acids Res ***42***, e155.

[47] Lebar, T., et al. (2014) *A bistable genetic switch based on designable DNA-binding domains*. Nat Commun ***5***, 5007.

[48] Kamihira, M., et al. (2005) *High-level expression of single-chain Fv-Fc fusion protein in serum and egg white of genetically manipulated chickens by using a retroviral vector*. J Virol ***79***, 10864-10874.

[49] Lenstra, J. A., et al. (1993) *Short interspersed nuclear element (SINE) sequences of the Bovidae*. Anim Genet ***24***, 33-39.

[50] Richardson, S. R., et al. (2015) *The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes*. Microbiol Spectr ***3***, MDNA3-0061-2014.

[51] Rangwala, S. H., and Kazazian, H. H., Jr. (2009) *The L1 retrotransposition assay: a retrospective and toolkit*. Methods ***49***, 219-226.

[52] Lavia, P., et al. (2022) *An Epigenetic LINE-1-Based Mechanism in Cancer*. Int J Mol Sci ***23***.

[53] Kawabe, Y., et al. (2018) *Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems*. J Biosci Bioeng ***125***, 599-605.

[54] Kawabe, Y., et al. (2017) *Improved recombinant antibody production by CHO cells using a production enhancer DNA element with repeated transgene integration at a predetermined chromosomal site*. J Biosci Bioeng ***123***, 390-397.

[55] Wang, X., et al. (2017) *Accumulative scFv-Fc antibody gene integration into the hprt chromosomal locus of Chinese hamster ovary cells*. J Biosci Bioeng ***124***, 583-590.

[56] Wang, T., et al. (2022) *Genetic switches based on nucleic acid strand displacement*. Curr Opin Biotechnol ***79***, 102867.

[57] Kawabe, Y., and Kamihira, M. (2022) *Novel cell lines derived from Chinese hamster kidney tissue*. PLoS One ***17***, e0266061.

[58] Qin, P., et al. (2022) *Switching the Activity of CRISPR/Cas12a Using an Allosteric Inhibitory Aptamer for Biosensing*. Anal Chem 10.1021/acs.analchem.2c04315.

[59] Falgenhauer, E., et al. (2022) *Transcriptional Interference in Toehold Switch-Based RNA Circuits*. ACS Synth Biol ***11***, 1735-1745.

[60] Halpern, M. E., et al. (2008) *Gal4/UAS transgenic tools and their application to zebrafish*. Zebrafish ***5***, 97-110.

[61] Siu, K. H., and Chen, W. (2019) *Riboregulated toehold-gated gRNA for programmable CRISPR-Cas9 function*. Nat Chem Biol ***15***, 217-220.

[62] Wang, S., et al. (2019) *A Novel Synthetic Toehold Switch for MicroRNA Detection in Mammalian Cells*. ACS Synth Biol ***8***, 1079-1088.

[63] Oesinghaus, L., and Simmel, F. C. (2021) *Controlling Gene Expression in Mammalian Cells Using Multiplexed Conditional Guide RNAs for Cas12a**. Angew Chem Int Ed Engl ***60***, 23894-23902.

[64] Perera, R. P., and Dosch, R. (2021) *In Vivo Imaging of Protein Interactions in the Germplasm with Bimolecular Fluorescent Complementation*. Methods Mol Biol ***2218***, 303-317.

[65] Schwarz, K. A., et al. (2017) *Rewiring human cellular input-output using modular extracellular sensors*. Nat Chem Biol ***13***, 202-209.

[66] DeRose, R., et al. (2013) *Manipulating signaling at will: chemically-inducible dimerization (CID) techniques resolve problems in cell biology*. Pflugers Arch ***465***, 409-417.

[67] Inobe, T., and Nukina, N. (2016) *Rapamycin-induced oligomer formation*

system of FRB-FKBP fusion proteins. J Biosci Bioeng ***122***, 40-46.

[68] Liang, F. S., et al. (2011) *Engineering the ABA plant stress pathway for regulation of induced proximity*. Sci Signal ***4***, rs2.

[69] Baeumler, T. A., et al. (2017) *Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors*. Cell Rep ***20***, 2639-2653.

[70] Tutucci, E., et al. (2018) *An improved MS2 system for accurate reporting of the mRNA life cycle*. Nat Methods ***15***, 81-89.

[71] Le, P., et al. (2022) *Illuminating RNA biology through imaging*. Nat Cell Biol ***24***, 815-824.

[72] Park, H. Y., et al. (2014) *Visualization of dynamics of single endogenous mRNA labeled in live mouse*. Science ***343***, 422-424.

[73] Nelles, D. A., et al. (2016) *Programmable RNA Tracking in Live Cells with CRISPR/Cas9*. Cell ***165***, 488-496.

[74] Fusco, D., et al. (2003) *Single mRNA molecules demonstrate probabilistic movement in living mammalian cells*. Curr Biol ***13***, 161-167.

[75] Konermann, S., et al. (2015) *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex*. Nature ***517***, 583-588.

[76] Ma, H., et al. (2018) *CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging*. Nat Methods ***15***, 928-931.

[77] Zadeh, J. N., et al. (2011) *NUPACK: Analysis and design of nucleic acid systems*. J Comput Chem ***32***, 170-173.

[78] Zuker, M. (2003) *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Res ***31***, 3406-3415.

[79] Wolfe, B. R., et al. (2017) *Constrained Multistate Sequence Design for Nucleic Acid Reaction Pathway Engineering*. J Am Chem Soc ***139***, 3134-3144.

[80] Fornace, M. E., et al. (2020) *A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed*. ACS Synth Biol ***9***, 2665-2678.

[81] Gossen, M., and Bujard, H. (1992) *Tight control of gene expression in mammalian cells by tetracycline-responsive promoters*. Proc Natl Acad Sci U S A ***89***, 5547-5551.

[82] Lim, F., and Peabody, D. S. (2002) *RNA recognition site of PP7 coat protein*. Nucleic Acids Res ***30***, 4138-4144.

[83] Romaniuk, P. J., et al. (1987) *RNA binding site of R17 coat protein*. Biochemistry ***26***, 1563-1568.

[84] Zhang, X., et al. (2010) *Conformational distributions at the N-peptide/boxB RNA interface studied using site-directed spin labeling*. RNA ***16***, 2474-2483.

[85] Hattman, S. (1999) *Unusual transcriptional and translational regulation of the bacteriophage Mu mom operon*. Pharmacol Ther ***84***, 367-388.

[86] Berens, C., and Hillen, W. (2004) *Gene regulation by tetracyclines*. Genet Eng (N Y) ***26***, 255-277.

[87] Zalatan, J. G., et al. (2015) *Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds*. Cell ***160***, 339-350.

[88] Hunt, V. M., and Chen, W. (2022) *Deciphering the Design Rules of Toehold-Gated sgRNA for Conditional Activation of Gene Expression and Protein Degradation in Mammalian Cells*. ACS Synth Biol ***11***, 397-405.

[89] Ebrahimkhani, M. R., and Ebisuya, M. (2019) *Synthetic developmental biology: build and control multicellular systems*. Curr Opin Chem Biol ***52***, 9-15.

[90] Sharker, S. M., and Rahman, A. (2021) *A Review on the Current Methods of Chinese Hamster Ovary (CHO) Cells Cultivation for the Production of Therapeutic Protein*. Curr Drug Discov Technol ***18***, 354-364.

[91] Kim, J. Y., et al. (2012) *CHO cells in biotechnology for production of recombinant proteins: current state and further potential*. Appl Microbiol Biotechnol ***93***, 917-930.

[92] Kameyama, Y., et al. (2010) *An accumulative site-specific gene integration system using Cre recombinase-mediated cassette exchange*. Biotechnol Bioeng ***105***, 1106-1114.

[93] Obayashi, H., et al. (2012) *Accumulative gene integration into a pre-determined site using Cre/loxP*. J Biosci Bioeng ***113***, 381-388.

[94] Kitano, H., et al. (2021) *Development of a genetically modified hepatoma cell line with heat-inducible high liver function*. Cytotechnology ***73***, 353-362.