Transgene integration into the ovalbumin locus of chicken cells using CRISPR/Cas9 system for transgenic chicken bioreactors

石, 銘

Transgene integration into the ovalbumin locus of chicken cells using CRISPR/Cas9 system for transgenic chicken bioreactors

SHI MING

December 2019

TABLE OF CONTENTS

Abstract

Transgenic chickens have been expected to be used as living bioreactors for the production of biopharmaceutical proteins. Development of a tissue-specific expression system of exogenous genes is a major concern in the construction of transgenic chicken bioreactors. For this purpose, it is important to develop a transgene integration method with high efficiency and specificity of expression. In recent years, the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system has attracted much attention as a versatile genome editing tool because of its ease of handling and simplicity. In this thesis study, the CRISPR/Cas9 system was applied to targeted knock-in of transgene into the ovalbumin (*OVA*) locus of chicken cells. The guide RNA sequences (gRNAs) of the CRISPR/Cas9 system against the *OVA* prompter were designed and evaluated for oviduct-specific expression of exogenous gene. A reporter gene expression cassette was integrated into the *OVA* locus of chicken cells using CRISPR/Cas9 system mediated by homology-independent targeted integration (HITI). For the knock-in cells, the transgene expression was successfully induced by activation of the endogenous *OVA* promoter using the CRISPR transactivation system. These results would contribute to studies in generating transgenic chicken bioreactors and activation of tissue-specific promoters for the production of pharmaceutical proteins.

Chapter 1

Introduction

The production of pharmaceutical and industrial recombinant proteins is needed to meet the demands of modern society. Recombinant proteins are an important tool to study biological processes and vital movement. The production of recombinant proteins requires an appropriate expression system. Currently, there are four major recombinant protein expression systems: *Escherichia coli*, *Pichia pastoris*, baculovirus/insect cell, and mammalian cells. The recent progress in gene editing technology developed for the chicken genome has enabled the production of recombination proteins in chicken, which is a powerful and cost-effective bioreactor. Pharmaceutical proteins can be produced in the laying hen through the expression of recombination proteins in the eggs. The chicken ovalbumin (*OVA*) promoter exhibits strong tissue-specific characteristics and over 50% egg protein can be expressed from this promoter.

The selection of an appropriate expression system is dependent on the recombinant protein characteristics, application purpose, and ability to produce the activated protein. Although the proteins can be rapidly expressed in the *E. coli* system, this system lacks the ability for posttranslational modifications (PTMs), which are observed in eukaryotes. The PTMs, such as SUMOylation, phosphorylation, palmitoylation, and glycosylation are essential for protein activity. In contrast to the *E. coli* expression system, some recombination protein expression systems enable optimal protein folding and PTMs. An appropriate expression system for the production of a recombinant protein must be selected based on the protein mass, disulfide bond number, type of PTM, recombinant protein purification method, and application of protein. The applications of recombinant proteins include structural studies, in vitro activity studies, antigens for antibody production, in vivo studies, and clinical studies. (Tab. 1-1)

1.1 Recombination protein expression system

1.1.1 *Escherichia coli*

The *E. coli* expression system has several characteristics, such as known genetic background, low cost, ease of operation, and high expression volume. *E. coli* is still the most widely used expression system for the production of recombinant proteins with high efficiency. Currently, many commercialized pharmaceutical proteins, such as G-CSF, interferon, and growth factor are produced in the *E. coli* system, which accounts for the production of one-third of all pharmaceutical proteins [1].

The *E. coli* expression system includes expression vector and host bacteria. In addition to the inserted gene fragments, the complete vector must also include the origin of replication, selective screening marker, promoter, and transcriptional terminator [2]. The promoter determines the transcription starting site and transcription efficiency, which is the key factor affecting the expression level of foreign genes [3]. The ideal promoter should be optimally regulated to reduce the toxicity of protein expression on the cells and to improve the expression level. Although several *E. coli* promoter systems are described, only a few promoters are commonly used for the recombinant protein expression. In the *E. coli* expression system, the production of recombinant protein can be increased and the negative effects of metabolism and product toxicity can be reduced using a suitable promoter. Currently, the recombinant proteins are expressed using the temperature-induced, isopropyl β-thiogalactopyranoside (IPTG)-induced, or nutrition-induced promoters. In addition to the commonly used promoters, such as $P_{L,}$ P_{R}, *Ptrp*, $P_{TAC,}$ and *Plac*, several efficient promoters, such as *T7*, *ara*, and *cadA* promoters are used for protein expression in *E. coli* [4]. The *E. coli lac* promoter is weak and does not promote efficient expression. Hence, it is not commonly used for the expression of recombinant protein. The *Tac* promoter is a strong promoter, which includes the -35 region of the *trp* promoter and -10 region of the *lac* promoter. The *Tac* promoter exhibits 5 times higher expression efficiency than the *lacUV5* promoter. However,

this promoter is also prone to miss expression and is toxic to the cells. The *T7* RNA polymerase system designed using the *T7* promoter can achieve high expression levels of thousands of homologous and heterologous proteins in *E. coli* BL21 (DE3). However, *E. coli* BL21 (DE3) does not exhibit high cell density [5]. The phage P_L promoter, a temperature-induced promoter, is used to transform the foreign proteins that are sensitive to a temperature range of 30–42°C. At 30°C, the promoter is inhibited by the repressor. At 42°C, the function of the repressor is attenuated, which induces the expression of foreign gene [6]. The *phoA* promoter, an alkaline phosphatase promoter, along with a signal peptide sequence of protein transport can guide foreign proteins into the periplasmic space. The *phoA* promoter is mild and can promote continuous expression. Additionally, the *phoA* promoter does not require additional inducers. The activation of the *phoA* promoter is inhibited in the presence of excessive phosphate in the medium. Conversely, phosphate starvation could activate the expression of exogenous proteins [7]. Another commonly used nutrient-induced promoter is the *araBAD* arabinose promoter, which is induced by arabinose. The *araBAD* expression system can be used with other promoters to regulate the expression of two or more recombinant proteins. However, the use of nutrient-induced promoter to express foreign proteins limits the composition of the culture medium [8].

1.1.2 *Pichia pastoris*

In the last decade, *Pichia pastoris* was rapidly developed as a eukaryotic expression system. The original host of this expression system is *Pichia pastoris* NRRLY11430, which was first discovered by Ogata in 1969. This strain is reported to exhibit rapid growth characteristics in the medium and utilize methanol as the sole carbon source [9]. Since the discovery of this strain, the potential of using *Pichia pastoris* to produce single-cell protein and then processing it into high-quality animal feed has gained the attention of many companies. Salk Institute of Biochemistry and Philip oil company have rapidly developed the *Pichia pastoris* expression system.

Salk Institute of Biochemistry has successfully isolated the alcohol oxidase gene (*AOX1*) promoter from the host bacteria and constructed the relevant vectors. Subsequently, the technical manual of *Pichia pastoris* gene operation was developed by Salk Institute of Biochemistry. Philip oil company has performed several studies and has experience in the field of single-cell protein production using this promoter. Furthermore, Salk Institute of Biochemistry successfully achieved the high-efficiency expression of the exogenous protein in *Pichia pastoris*. In 1993, Philip oil company sold the patent of *Pichia pastoris* expression system to Research Corporation Technologies and entrusted Invitrogen company to sell the related products. Currently, there are several commercial *Pichia* expression system kits available in the market, which has enabled the easy expression of the exogenous protein in *Pichia pastoris*. Additionally, the availability of these kits has markedly increased the utilization of *Pichia pastoris* as an expression system in the past decade. Furthermore, several structural genome projects have used *Pichia pastoris* as the protein expression system for protein production platforms.

There are three types of *Pichia pastoris* based on the presence of screening markers: wild type, nutritional deficiency type, and protease deficiency type. The wild type strains include bg10 (from strain nrrly-11430) and X-33. The X-33 strain is generated using the histidine dehydrogenase deficient strain, GS115 through overexpression of histidine HIS4. The nutritional deficiency strains are mainly used to screen the recombinant transformants, such as the deletion mutants of HIS4, arg4, ade1, URA3, or their combination by comparing them to the wild type strains. There are three types [10] of protease deficient strains: smd1163, smd1165, and smd1168. Among these strains, smd1168 is the most commonly used strain for recombinant protein expression. Smd1168 (the carboxypeptidase pep4 (pep4: URA3) mutated on the basis of GS115 strain) can inhibit the hydrolysis of the carboxyl end of the recombinant protein (mainly the C-end of protein containing Lys and Arg basic amino acids) [11]. The main disadvantages of smd1168 are that this strain exhibits slow growth and low transformation efficiency when compared to other strains, such as

GS115 and X-33 strains. Various proteins exhibit different expression and degradation levels in *Pichia pastoris*. Hence, the *Pichia pastoris* strains should be selected according to different conditions.

Based on the speed of methanol metabolism, *Pichia pastoris* is divided into three types: Mut^+, Mut^S, and Mut^-. *Pichia pastoris* can utilize methanol as a carbon source, which is mainly because of the presence of alcohol oxidase (AOX) in the peroxisome. AOX can metabolize methanol to produce formaldehyde and hydrogen peroxide. Hydrogen peroxide is reduced to water and oxygen by catalase, while formaldehyde is converted into formic acid and carbon dioxide by formaldehyde dehydrogenase (FLD) and formic dehydrogenase (FDH), respectively. During this process, NAD is reduced to NADH and a large amount of ATP is produced through electron transfer and oxidative phosphorylation [12]. Therefore, the final metabolites of methanol can provide energy and carbon sources for yeast growth. Most AOX is encoded by the *AOX1* gene, while a few AOXs is encoded by the *AOX2* gene. Although the *AOX1* and *AOX2* gene sequences share 97% homology, their promoter sequence exhibit variation. The *AOX1* promoter promotes the expression of *AOX2* gene, while the *AOX2* promoter downregulates the expression of *AOX1* gene [13]. This indicated that the difference in the expression of *AOX1* and *AOX2* is mainly due to the promoter. Both *AOX1* and *AOX2* genes are expressed in the Mut^+ strain, which exhibits rapid growth in the methanol medium. The *AOX1* gene is nonfunctional in the Mut^S strain. If the *AOX1* gene in the KM71 strain is replaced by aox1::ARG4 sequence, AOX is expressed only by the *AOX2* gene. Additionally, this strain exhibits slow methanol metabolism and cell growth in the methanol medium. The Mut^- strain does not have functional *AOX1* and *AOX2* genes and thus cannot express AOX. Therefore, this strain is rarely used to express foreign proteins. The homologous recombination of plasmids determines the recombinant GS115 strain phenotype. If an insertion introduced in the linearized plasmids via homologous recombination does not disrupt the *AOX1* gene, the recombinant strain exhibits Mut^+ phenotype. If a substitution is introduced via homologous recombination, the transformant exhibits Mut^S phenotype.

1.1.3 Baculovirus/Insect Cells

Baculovirus is an encapsulated double-stranded circular DNA virus. The viral body is rod-shaped and thus named as baculovirus [14]. Baculovirus is mainly found in insects. Among the baculovriuses, Autographa calofornica nuclear polyhedrosis virus (AcMNPV) has been extensively studied. The expression system of autumn slime insect cell (*Spodoptera frugiperda*, SF) was first established by Smith using the strong polyhedrin gene promoter. The characteristics of large-scale expression in the late stage were successfully used to express human interferon β [15]. Baculovirus has been used as a vector to express foreign genes in insect cells or body. Baculovirus expression vector system (BEVS) is one of the four major expression systems in the field of genetic engineering. BEVS has increasingly important applications in studies on gene expression regulation, protein structure and function analysis, as well as the production of various bioactive substances [16].

Baculoviruses can be divided into two subfamilies: Baculoviridae (inclusion body baculoviridae) and nudibaculovirinae (non-inclusion body baculovirinae). The inclusion body baculoviruses can be divided into nuclear polyhedrosis virus (NPV) and granulosis virus (GV). There are several NPV particles in the protein crystal of the single nucleus. According to the degree of its core-shell aggregation, NPV can be further divided into single entrapment (SNPV) and multiple entrapment types (MNPV) (Fig. 1-1). Generally, there is only one virus particle in the protein crystal of GV. The non-occluded baculovirus does not form the inclusion body, which may be due to the lack of gene encoding crystal protein.

Insect baculovirus is an enveloped closed double-stranded DNA virus. The circular DNA has an average length of 135 kb (80–180 kb) [16]. AcMNPV C6 strain was the first baculovirus whose whole genome was completely sequenced. The genome length of AcMNPV C6 strain is 133894 bp, which includes 59% A+ T content and 337 open reading frames (ORFs) (with a length higher than 150 bp) that are evenly distributed in two chains of the whole genome [17]. Ahrens et al. reported the complete nucleotide sequence of Orgyia pseudotsugata multinucleocapsid NPV

(OpMNPV) of the yellow cedar moth in 1997 [18]. The sequence analysis of BmNPV which from Bombyxmori ovaries revealed that the genome length was 128413 bp, which comprises 40% G + C content and 136 ORFs that are predicted to encode more than 60 amino acids. The ORF amino acid sequence homology of BmNPV to that of AcMNPV was about 90% [19]. The total length of LdMNPV is 161046 bp, which comprises 57.5% G + C content and 163 estimated ORFs [20]. Recently, various NPV genomes have been sequenced [21, 22].

Based on the time of gene expression relative to DNA replication of baculovirus, the virus genes are divided into two comprehensive phases. The genes expressed before the initiation of DNA replication are called early genes, while those genes expressed at the beginning of DNA replication or later are called late genes. During the late stage, polyhedrin and P10 proteins are highly expressed. Polyhedrin is the main component for the formation of inclusion bodies. The accumulation of polyhedrin in cells can be as high as 30–50% during the later stages of infection. Although polyhedron is not necessary for virus replication, it exerts a protective effect on the virus particles, which can keep them stable and infectious. P10 protein is also a non-essential component of virus replication. P10 can form fibrous substances in cells, which may be related to cell lysis. The polyhedral and P10 genes have been mapped and cloned. These two genes have strong promoters and thus these two gene loci are the ideal foreign gene insertion sites of the BEVS [23].

Baculovirus genes are copied and transcribed in the insect nucleus [16]. The DNA is replicated and assembled in the baculovirus nucleocapsid, which is flexible and can accommodate larger fragments of foreign DNA. Thus, baculovirus is an ideal vector for the expression of large fragments of DNA. Among the baculoviruses, only NPV is used as a foreign gene expression vector. The virus particles can be packaged in polyhedrin to form a 1–5 μm long inclusion body virus in the shape of polyhedron [24]. There are two forms of NPV [25]: occlusion derived virus (ODV) and extracellular budding virus (BV). Although these two forms have the same genetic information [26], they use different ways to infect the host. The inclusion body virus

is horizontally transmitted in the insects, which often causes infection after the host consumes food contaminated with ODV. The inclusion body virus is coated with a layer of protein crystal that comprises around 29,000 polyhedral proteins, which play an important role in the horizontal transmission of the virus. The polyhedral proteins protect the virus particles from inactivation by environmental factors during the process of external transmission. Additionally, the polyhedral proteins ensure that the virus particles are released at appropriate locations to cause infection. The strong alkaline environment (pH = 10.5) in the local midgut epithelium of insects can induce the virus particles to release protease to dissolve the polyhedron [27]. BV, which is produced by cell buds, is transmitted between cells. BV then enters the hemolytic system and infects the cells in other regions or directly infects the surrounding cells.

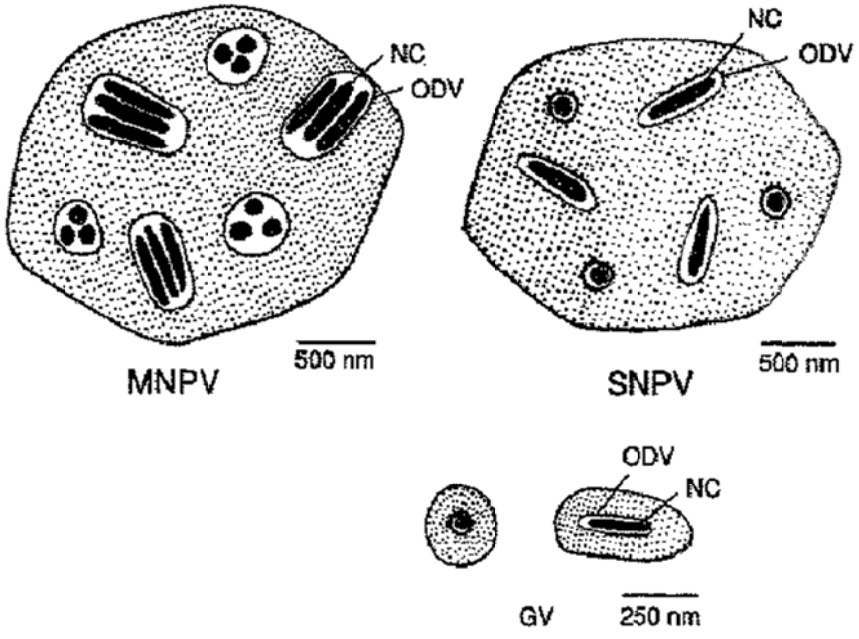


Fig. 1-1 Schematic representation of structure of baculoviridae [16]

1.1.4 Mammalian Cells

The most commonly used mammalian cell expression systems are CHO, NS0, HEK293, COS, BHK, and SP2/0. Among them, CHO, NS0, and HEK293 are the three most important and commonly used host cells [28]. CHO and NS0 are mostly used to construct stable expression cell lines, which can be used for commercial

production. The NS0 cells are mainly used in the production of recombinant therapeutic antibodies, while the CHO cells are not only used for the production of recombinant therapeutic antibodies but also for the expression of other recombinant proteins. The HEK293 cells are mainly used for transient expression as they have high transfection efficiency.

The human embryonic kidney cells (HEK293) are immortalized by transfecting adenovirus type 5 (Ad5) DNA into the primary human embryonic kidney cells. Since the generation of these cells in 1977, several derived cells have been widely used for transient expression [29, 30]. The 293-F cells in suspension culture are domesticated by wild type HEK293 cells with good growth status and high expression of exogenous protein [31]. Other derived 293 cells have been genetically engineered, such as 293EBNA cells that stably express Epstein Barr virus nuclear antigen (EBNA-1), 293T cells that stably express SV40 virus large T antigen, and HKB-11 cells that are a fusion of 293 cells and B lymphocytes. Although the protein expression in the HEK293 derived cells is slightly different during transient expression, these cells can be easily transfected compared to other mammalian cells. Therefore, the 293 cells have been the preferred host cells for transient gene expression. The titer of the recombinant antibody after instantaneous expression in the 293E cells can reach 1 g/L, which was reported as early as 2008 [32].

Although a large number of mammalian cell lines can be selected to express foreign proteins, almost 70% of the recombinant proteins are obtained by stable expression in the recombinant CHO cell lines [33]. The wide application of CHO cells can be attributed to the following advantages: (1) well-known genetic background and stable physiological metabolism, which are recognized as safe gene engineering receptor cells by the FDA; (2) CHOs are fibroblasts cells, which rarely secrete their own internal proteins and are conducive to the separation and purification of external proteins; (3) Accurate modification of external proteins; (4) Optimal gene transfer and expression system; (5) Exhibit good shear resistance and is easy to culture in a serum-free medium with high density and large-scale suspension; (6) After the foreign

gene is integrated into the CHO cell chromosome, it can be maintained stably without selection pressure. The main disadvantage of the CHO cell expression system is the low yield. An important way to improve the expression level of foreign genes in the CHO cell is to use the gene screening amplification system to build a stable and high-yield CHO cell line. This mainly includes two kinds of gene amplification mediated by dihydrofolate reductase (DHFR) and glutamine synthetase (GS) system.

After the introduction of foreign target genes into the CHO cells, the foreign genes integrate into the genome of only a few cells. These foreign genes can be transcribed and expressed for long-term and the stable expression of the target protein can be maintained. To construct a stable CHO cell line, selective markers must be used for screening. The cells that contain the integrated foreign genes in the genome can be screened out by co-expressing the selective markers with the target gene. The selective markers can be divided into two types: non-amplified genes, such as neomycin, which has no effect on the copy number of the target gene and amplified genes, such as the DHFR and GS genes. A high copy number of the target gene can be obtained by enhanced expression of target protein, which is an important way to improve the expression level of foreign genes in the CHO cells. Currently, almost all recombinant proteins used in commercial production are stably expressed using this gene amplification strategy. The recombinant CHO cell line is a commonly used platform for the production of recombinant therapeutic proteins in the biopharmaceutical industry. The most commonly used gene amplification system is the DHFR system [33]. DHFR can be inhibited by methotrexate (MTX), a folate analog. When the expression plasmids carrying the *DHFR* gene and target gene are transfected into the *DHFR* gene-deficient CHO cells or wild type CHO cells, the clones can be obtained by culturing the cells in the selective medium. The selection pressure of MTX resistance promotes the co-expression of foreign genes and DHFR. The copy number can increase hundreds to thousands of times, which results in high target gene expression [34].

The selection pressure markedly increases the target gene expression and the

stability of the cell line. It is beneficial to obtain high-yield recombinant cell lines and stable target gene expression in the cell lines. High expression clone cells can be obtained in a short period of time by large-scale and rapid selective pressure. However, the final expression level obtained using the large-scale and rapid pressure strategy will not be higher than that obtained using the small-scale and repeated pressure strategy. The target gene expression level in the cell lines is unstable and often drops after the removal of selection pressure. The limitation of the DHFR amplification system is mainly associated with the time-consuming and laborious multi-level repeated screening of resistant cells. The whole screening process usually lasts for 4–6 months. Additionally, the gene amplification is unstable after the removal of the selection pressure. The selection pressure may result in complex composition and poor product homogeneity because the gene amplification range is not fixed.

The GS system is another effective gene amplification system that has been successfully applied in the CHO cells expressing GS [35, 36]. The CHO cells contain endogenous *GS* gene and hence the GS inhibitor, methionine sulfoximine (MSX) inhibits the endogenous GS activity. Similar to the MTX screening process, the expression of *GS* gene and its related target gene increases with MSX concentration, which results in enhanced target gene expression level. The number of foreign gene copies of the recombinant cell lines increased by more than 200 times. The main advantage of this system is that there is no need to add glutamine in the cell culture medium. Hence, toxic ammonia will not accumulate during culturing. The screening intensity and GS efficiency have greatly improved with the development of a GS knockout CHO cell line (CHOK1SV). The major limitation of the GS system is that the long-term continuous culture results in poor growth conditions for the cells.

Currently, the foreign genes are mainly integrated into the host cell genome randomly, which usually requires multiple rounds of cloning and screening to produce acceptable high expression clones. However, re-screening the monoclonal cells is necessary whenever a new expression system is constructed. The homologous recombination or clustered regularly interspaced short palindromic repeats

(CRISPR)/CRISPR-associated protein 9 (Cas9)-mediated homologous recombination technology can achieve homologous recombination at known transcriptionally strong active regions and form "landing pad" that can insert the target genes at specific points, most of which are in the strong transcriptional active regions. The target gene can be expressed efficiently and stably with or without selective pressure [28].

Table 1-1 Summary of expression methods [37]

Expression systems	Advantages	Disadvantages
E. coli	Rapid expression method (days)	Limited capacity for posttranslational modifications
	Inexpensive bioproduction media and high density biomass	Difficult to produce some proteins in a soluble, properly folded state
	Simple process scale-up	
	Well characterized genetics	
P. pastoris	Moderately rapid expression method (weeks)	N-linked glycan structures different from mammalian forms
	Inexpensive bioproduction media and high density biomass	Enhanced safety precautions needed for large-scale bioproduction due to methanol in induction media
	Most posttranslational modifications and high folding capacity	
Baculovirus/ insect cell	Moderately rapid expression method (weeks)	N-linked glycan structures different from mammalian forms
	Most posttranslational modifications and high folding capacity	Low density biomass and expensive bioproduction media
		Difficult process scale-up
Mammalian—transient expression	Moderately rapid expression method (weeks)	Low density biomass and expensive bioproduction media
	All posttranslational modifications and high folding capacity	Difficult process scale-up
Mammalian—stable expression	All posttranslational modifications and high folding capacity	Lengthy expression method (months)
		Low density biomass and expensive bioproduction media
		Difficult process scale-up

1.2 Transgenic animal bioreactor

The modern molecular genetic technologies along with cell engineering and embryo engineering are used to modify the animal genome to produce some unnatural

characters or to improve the production performance. One of the most attractive prospects of transgenic technology is that transgenic animals are transformed into efficient biopharmaceutical "factory," which can rapidly produce cytokines, serum proteins, and recombinant antigens for immunotherapy and some antibodies for passive immunotherapy.

There are two important breakthroughs in biological research that have enabled the construction of the transgenic animal bioreactor: generation of transgenic mice and the discovery of tissue-specific expression regulatory elements. Godrno et al. [38, 39] injected DNA into the pronucleus of the mouse fertilized eggs by microinjection, which demonstrated that the foreign genes can be integrated into the genome of recipient mice and can be transmitted to the offspring. The tissue-specific expression regulatory elements have enabled the foreign genes to be localized and expressed in specific tissues [40-42]. The pioneering work of Clark et al. resulted in transgenic mice exhibiting mammary gland-specific expression of foreign genes, which enabled the modification of the nutritional composition of animal milk or "human emulsification" of animal milk [43-45]. Subsequently, the transgenic cows [46], goats [47], sheep [47] and pigs [48] expressing the recombinant protein in the mammary gland were generated successively. These breakthrough studies on transgenic animal bioreactors have resulted in increased investment in technology-related research by biotechnology companies.

1.3 Transgenic chicken oviduct bioreactor

The main advantages of transgenic animal mammary gland bioreactor include high expression quantity and complete activity of expression products, but there are some disadvantages such as long research cycle, long generation interval and high development cost, while transgenic chicken bioreactor can not only avoid the above problems, but also has some outstanding advantages. Each egg contains about 3.5–4.0 g protein, which ovalbumin (> 50%) is the main protein. The expression level of other four proteins (lysozyme, ovomucoid, ovomucin and companion protein) is about 50

mg. More than 300 eggs can be laid each year in excellent breed of laying hens. If each egg produces 1g of medicinal protein, 3–4 chickens can produce 1kg of protein in one year. Moreover, the protein composition in the egg white is relatively simple, and the purification of recombinant protein is relatively convenient. Moreover, there has been successful experience in extracting the target egg white protein such as lysozyme from the egg. The egg white contains natural protease inhibitor, plus natural sterile micro environment provides a guarantee for avoiding the pollution of recombinant protein and ensuring its biological activity stability [49]. Compared with mammalian, the glycosylation pattern of chicken derived protein is more similar to that of human derived protein [50]. Therefore, chicken is very promising as a low-cost and high-yield bioreactor, but its special reproductive biological characteristics make the research of transgenic chicken far behind other animals.

1.3.1 Basic principle of chicken oviduct bioreactor

The basic principle of chicken oviduct bioreactor is to use the regulatory sequence in the chicken gene for oviduct-specific expression of foreign genes and secrete the expression products into the egg white. The main techniques used in this bioreactor include the construction of oviduct-specific expression vector, gene transfection into the chicken fertilized eggs (embryos), and breeding of transgenic chicken offspring. The construction of chicken oviduct-specific expression vector and the tissue specificity of driving foreign gene expression are keys to develop transgenic chicken bioreactors.

1.3.2 Fertilization and early development of chicken embryo

There is little progress in poultry transgenic research when compared to other animals because of the complex physiological process from hatching to producing fertilized eggs. The production of eggs starts from the ovulation of mature eggs. The yolk is equivalent to the oocyte. The sperm enters the vagina and some of them are stored in the seminal gland (the special tube gland of the shell gland and the vagina

junction). The sperms are then continuously transported to the funnel of the fallopian tube. The egg enters the funnel after ovulation and fertilizes with the sperm. The fertilized egg enters the enlarged part of the fallopian tube, where it stays for at least 2.5 h and the gland is formed. The fusion of female and male pronucleus usually happened within 3–4 h after ovulation [51, 52]. The pronucleus is located at 22–52 μm below the yolk membrane [52], which cannot be observed without fixation. Hence, the formation of eggs takes at least 22 h and the early development of embryos is initiated in the reproductive system of female birds [53]. Therefore, even the newly produced eggs have developed into a disc (located in the blastocyst cavity) with about 60,000 undifferentiated totipotent cells (Fig 1-2). The fertilized egg and early embryo have huge yolk, which prevents direct gene transfection. In the newly produced egg, the embryo develops to the late stage of the blastocyst (named as the tenth stage, i.e. the X stage). The blastoderm consists of 60,000 pluripotent blastomeres with clear demarcation between blastocyst cavity and yolk. Next, the primordial germ cells begin to appear. Once the egg is produced, the embryo stops developing. The embryo development will restart when suitable hatching conditions are provided. After hatching for 18 h, the primordial germ cells are formed and migrate to the new moon. Next, the primordial germ cells enter the blood system and finally the developing gonads.

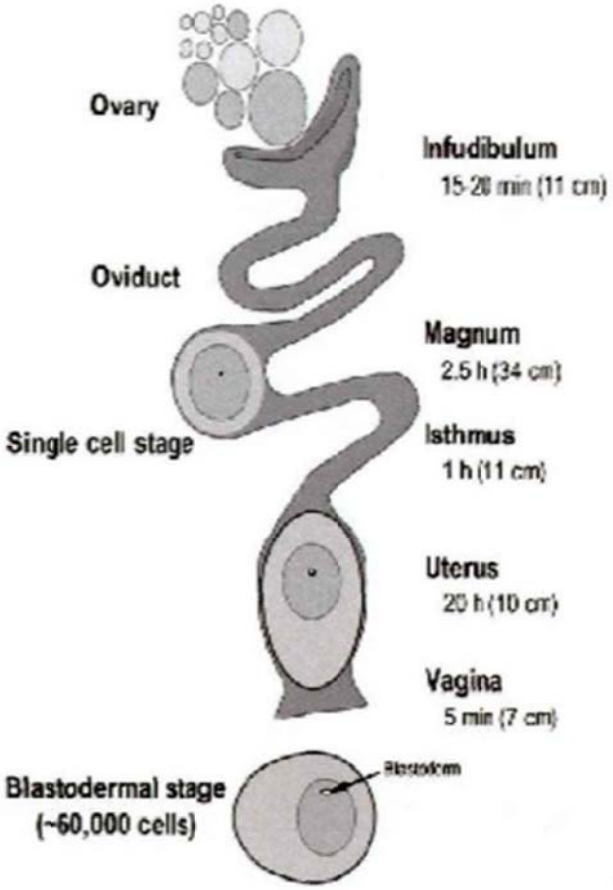


Fig 1-2 Egg information in hen [54]

1.3.3 Ovalbumin

Ovalbumin is specifically expressed in the chicken oviduct. Ovalbumin is secreted into the lumen of oviduct by the secretory cells of oviduct epithelium through acroplasmic secretion. Ovalbumin, a phosphoglycoprotein, is made up of 386 amino acid residues and has a molecular weight of 43 kDa. The length of the ovalbumin coding gene is 760 bp, while the length of the mature mRNA is 1873 nucleotides.

The ovalbumin gene belongs to the same gene group as ovotransferrin, ovomucoid, and lysozyme genes, which are located on the long arm of chromosome 2. The ovalbumin gene family consists of x gene, y gene, and ovalbumin gene [55], which are arranged sequentially from 5′ to 3′ direction. In the chicken somatic cells, each haploid genome has only one copy of the ovalbumin gene, which comprises 8 exons and 7 introns. The length of the first intron is 1.5 kb. The noncoding region of

the 5' end is divided into two parts: leading region (47 bp) and noncoding region (17 bp). The last exon of the ovalbumin gene is one of the largest known exons, accounting for more than half of the whole mRNA coding region, with a total length of 1043 bp.

The chicken ovalbumin gene has regions that are associated with transcriptional activity, such as DNase I hypersensitive region. The ovalbumin gene and β-globin gene extracted from the chicken oviduct cells are degraded by DNase I. There are four hormone-induced DNase I hypersensitive sites [56, 57] in the 5' terminal regulatory region of the ovalbumin gene. The transcription factors easily bind to these sites, allowing the rotation of DNA. Site I is a group of negative regulatory elements (NREs) located between -308 and -88 regions, including -308 and -256 region, -239 and -220 region, and -174 and -88 region. Some oligomers corresponding to these three regulatory elements can also inhibit the promoter of the thymidine kinase gene. Several studies have demonstrated that the region between -239 and -220 is a transcriptional silencer and its sequence (TCTCTCCNA) is consistent with that in other genes. The oviduct-specific proteins can bind to the following regions: -280 and -252 region and -134 and -88 region. These regions may be potential tissue-specific expression regulators of the ovalbumin gene [58]. Site II is a steroid-dependent regulatory element (SDRE) located between the -892 and -780 region, which can combine with estrogen and glucocorticoid to produce the corresponding response. Additionally, the SDRE is an important regulatory element for hormone-induced gene expression in the chicken oviduct cells in vitro. NRE is reported to exert dual effects. In the absence of steroids, NRE can inhibit the ovalbumin gene transcription. In the presence of steroids, NRE and SDRE can promote the ovalbumin gene transcription [59]. DNase I hypersensitive site contains the direct repeat sequence of estrogen semi responsive element with palindrome structure, which can mediate hormone-induced gene expression in the HeLa cells [60] and enhance the transcription activity of the upstream chicken ovalbumin gene promoter [61]. The function of the remote DNase I sensitive point IV is still unknown.

The whole ovalbumin gene family is distributed in the 120 kb region [62]. Thus, there may be other cis-regulatory elements [63, 64] besides the ovalbumin gene. Snadesr et al. demonstrated that the downstream sequence of ovalbumin gene (-880) was sufficient to promote high steroid-induced reporter gene expression level, which was lower than the level guided by the regulatory regions of -2.8 kb and -3.5 kb [59]. There may be tissue-specific expression regulators in the region between -3200 and -2800, which inhibit the expression of exogenous genes in the liver driven by ovalbumin gene regulatory region [65].

The mRNA structure of the ovalbumin gene has been elucidated. There is a hairpin structure at the 5' end of the ovalbumin gene, which has a high affinity for the eukaryotic initiation factor, eIF-2. It can be used as an initiation signal to facilitate the translation initiation of mRNA [66]. Several studies have demonstrated that steroids can enhance the ovalbumin gene transcription and contribute to the stability of mRNA [67].

1.4 Gene editing

There is a renewed interest in gene research since the discovery of genes as the carrier of genetic information and the elucidation of DNA structure. The development of new sequencing technology has enabled the sequencing of the whole genome of several species. The sequencing data can aid in understanding the application of gene function. The manipulation of the genome is called “genome editing,” which involves deletion, insertion or substitution of genomic DNA sections using the DNA repair mechanism in the cell. The traditional genome editing is mainly used to modify the genome by spontaneous homologous recombination and integration of foreign genes. However, this method is very inefficient (about 10^{-7}) and is associated with many random integrations [68]. Hence, traditional genome editing is only suitable for screening rare targeted yeast cells and cultivable mammalian cells by detecting a large number of transfected cells [69, 70].

DNA damage can produce DNA double-strand breaks (DSBs), which are usually

toxic to cells and results in genomic instability and gene mutations that cause diseases. However, the efficiency of gene targeting increased by 50,000 times when the target site has DSBs [71]. DSBs stimulate the DNA repair mechanism *in vivo*. The DNA repair mechanism can not only repair the damage but is also involved in some important biological processes, such as meiotic recombination, antibody class conversion, and VDJ rearrangement. DNA DSBs induce two main repair mechanisms: non-homologous end joining (NHEJ) and homology directed repair (HDR). HDR uses homologous DNA sequence as a template to repair the cleaved DNA double strand and the damaged DNA sequence is strictly complementary to the template DNA. This repair is catalyzed by RecA protein in bacteria and Rad51 protein in eukaryotes [72]. Homologous recombination usually occurs in the S/G2 phase of cell cycle. In the absence of homologous sequences, DSBs are repaired by NHEJ. This mechanism can cause mismatches, which can easily change the sequence of genes and result in deletion or insertion of small fragments of bases at the damaged sites. NHEJ is active during the whole cell cycle. Although NHEJ is prone to mismatch, this is the main repair mechanism for DSBs [73]. NHEJ provides an effective way to interrupt gene function by knocking out the gene. HDR provides another editing method, which relies on the knock-in of homologous sequence, by precisely copying the recombinant genetic information on the template into the repaired DNA. NHEJ and HDR are two important ways of gene editing in almost all living organisms. Therefore, these two methods can be theoretically be applied to any species [74].

As HDR repairs the damaged DNA in strict accordance with the donor homologous template, researchers used HDR to insert and repair mammalian cell genes in the 1980s [75]. This technology was then used by Capecchi to target the mouse embryonic stem cells and successfully obtained transgenic mice [68]. However, this technology can only produce effective HDR in yeast, chicken DT40 cells, and mouse embryonic stem cells among eukaryotes [76]. This traditional DNA homologous recombination technology is time-consuming, laborious, and inefficient, which hinders the wide application of gene editing technology. Thus, researchers

searched for an endonuclease that can cut the genomic sites at specific sites and produce DSBs. The first nuclease used for genome editing of mammalian cells is the yeast homing endonuclease, I-SceI. I-SceI recognizes a special 18-bp sequence of chromosomes. This endonuclease exhibits high specificity as the recognition sites are not present in the mouse and human genomes. DSBs are produced by I-SceI cleavage, which improves the HDR efficiency of mammalian cell genome target sites by 500 times [77]. As the recognition sites of I-SceI endonuclease are present only in yeast, it cannot be used to edit endogenous genes in mammalian genome, which limits its application. Studies on I-SceI indicate that designing site-specific endonucleases will promote the development of genome editing technology.

One of the challenges of genome editing is to identify a nuclease that can specifically recognize and cut the genome to produce DSBs [78]. Recently, researchers have realized that a wider range of nuclease target DNA sequence combinations can be employed [79]. Based on this concept, the following two nucleases have been artificially developed: zinc finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN). ZFN and TALEN both cleave DNA via the FokI domain. In contrast to ZFN and TALEN, CRISPR/Cas, which is a third-generation artificial nuclease, is an RNA-mediated genome editing technology. CRISPR/Cas system is a widely used genome editing tool mainly due to its simplicity, low consumption, and high efficiency.

The emergence of endonuclease that can specifically cleave the genomic site has enabled accurate editing of the target gene sites. The development of genomic editing technology promotes innovation in the field of gene research. Researchers can now use artificial nuclease to delete or integrate any gene [80]. The new genome editing technology is accurate as it can modify the genome by targeting the gene sequence. Thus, the new genome editing technologies have more advantages than traditional transgenic methods and RNAi-mediated knockdown technology. Artificial nuclease technology can not only modify the coding gene but also edit the cis-regulatory elements, which results in gain or loss of gene function. Hence, artificial nuclease

technology can sensitively and accurately reflect the expression and function of endogenous genes.

1.4.1 ZFNs

ZFN is a chimeric protein comprising a zinc finger module and a FokI nuclease domain [79]. The single nuclease domain lacks the specificity of DNA recognition and relies on the fused zinc finger protein to specifically recognize the DNA sequence. The zinc finger domain is a natural DNA binding domain, which is present in transcription factors. One zinc finger module can recognize three nucleotide bases. The recognition of DNA sequences by protein domains, such as zinc finger allows researchers to assemble an artificial DNA binding module based on zinc finger. The first generation of tools that was directly used for genome editing comprised designable DNA binding sites and nuclease domain [81]. Type II restriction endonucleases, such as FokI can recognize short DNA sequences and induce DSBs by cutting the DNA at a certain distance between recognition sites. This is because the DNA binding sites and nuclease domains of these proteins are separate and can function independently [82]. The nuclease domain does not exhibit sequence specificity and the site of DNA break can be changed by changing the specificity of DNA binding domain [83]. Cys2-His2 family is the most suitable zinc finger protein for application in combination with fixed-point nuclease. Each zinc finger is very small with about 30 amino acids. The secondary structure consists of an α helix and two β sheets (αββ). One of the characteristics of zinc finger endonuclease is that FokI needs to form dimer to exhibit the enzyme activity of cutting DNA double strand [84]. Therefore, the two zinc finger proteins must be combined with FokI, which allows FokI to form a dimer and cleave the DNA double strand. The formation of nuclease domain dimer increases the recognition accuracy as the length of the recognition sequence doubles. In the DNA binding domain and target sequence, two nuclease domains must be engineered close to each other to form dimer for cleaving the DNA. The best distance between two zinc finger domains is 5–7 nucleotides [85].

Since ZFN was first successfully used to knockout the yellow gene of Drosophila in 2002, it has been successfully used to modify different genes of Drosophila, zebrafish, mouse, rat, pig, silkworm, Arabidopsis, and humans [86-91]. The efficiency of different gene editors varies with an average efficiency of about 10% [81]. Although the ZFN-mediated genome editing method is more efficient than the traditional methods for site-specific integration, it is a time-consuming process to edit genes with ZFN. To select a specific target site, researchers must build a zinc finger expression library and many zinc finger proteins do not exhibit activity. The cost and time for screening high efficiency and specific zinc finger proteins increases.

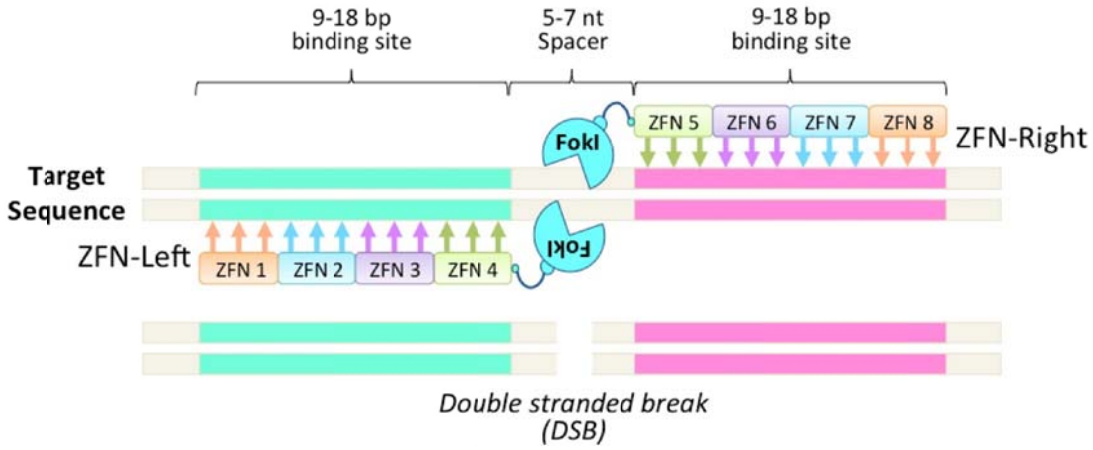


Fig. 1-3 Schematic diagram of ZFN components and genome-editing processes.

1.4.3 TALEN

Transcription activator-like effector (TALE) is a secreted protein from *Xanthomonas*, a plant pathogenic bacterium. TALE protein enters the plant cells through type III secretion system. Next, TALE enters the nucleus to activate the expression of downstream target genes, which aids in virus proliferation and transmission [92, 93]. Typical TALE proteins contain some identical regions with certain structural characteristics, including nuclear localization signal, N-terminal transport signal, central repeat DNA binding region, and C-terminal transcriptional activation region (Fig 1-4). A typical TALE DNA binding region consists of 15.5–19.5 tandem repeats, each of which contains 34 highly conserved amino acid residues. Among the conserved amino acid residues, the twelfth and thirteenth amino acids are

called repeat variable residues (RVDS), which determine the specificity of DNA binding region for a nucleotide. The last repeat unit has only 20 amino acids, which is only half a repeat. Bioinformatic laboratory studies have demonstrated that there is a certain correspondence between RVD of TALE and nucleotide of the target DNA sequence. RVD with different amino acid combinations can recognize one or more of A, T, G, and C bases. Common corresponding relationships are as follows: histidine aspartic acid (HD) recognizes base C; asparagine isoleucine (Ni) recognizes base A; asparagine asparagine (NN) recognizes base A or G; asparagine glycine (NG) recognizes base T; asparagine serine (NS) can recognize any of A, T, G, C; asparagine lysine (NK) recognizes base G [94]. After the discovery of TALE code, TALEN, another landmark gene editing tool, was developed by combining TALE protein with nuclease. Similar to ZFN, a pair of two TALE proteins bind to the fusion protein of nucleic acid endonuclease *Fok*I to form a dimer. Generally, the distance between two TALENs and the corresponding genomic DNA binding sites is 10–20 bp for cleaving the specific target sites on the target genome. Some different assembly methods are used to generate customized TALENs. The TAL module area and simple DNA recognition code enable us to assemble TALENs that easily and rapidly target any gene of interest. Generally, TALEN can recognize 18–20 bp DNA sequences and increasing the number of DNA binding regions on TALENs will reduce the specificity [95]. The characteristics of the TAL region in TALEN enable the prediction of target specificity. Since the advent of TALEN technology, genome modification has been successfully accomplished in zebrafish, mouse, fruit fly, rat, frog, human cell, insect, and various plants in a short time [96-99]. Although the size of TALE protein is similar to zinc finger protein, the TALE protein can recognize only one base. Thus, the final constructed nuclease could be larger. Additionally, the difficulty of transfection is also a limitation. Similar to other nucleases, TALEN can miss the target sequence in the genome. This can be addressed by selecting a unique site (at least 7 nucleotides) that is different from other sites in the genome.

Compared to ZFN, the construction of TALEN is easier. TALEN can complete

the screening with simple molecular cloning technology, which does is not time-consuming. Compared to ZFN, TALEN has a wider target site selection range, stronger ability to recognize the specific DNA sequence, better specificity, and higher target efficiency. However, TALEN technology is similar to ZFN. Only 34 amino acids of a TAL module can recognize a base. Therefore, to build a TALEN protein that can recognize the specific sequence of a target DNA, it may be necessary to build a TALEN protein with hundreds of amino acids. In addition to the time-consuming assembly, TALEN protein may also elicit an immune response in vivo and reduce the editing efficiency. Before TALEN could be widely used as an alternative artificial nuclease to ZFN, the RNA-mediated nuclease, CRISPR/Cas system emerged. This new technology directly replaced TALEN to be the predominant genome editing technology.

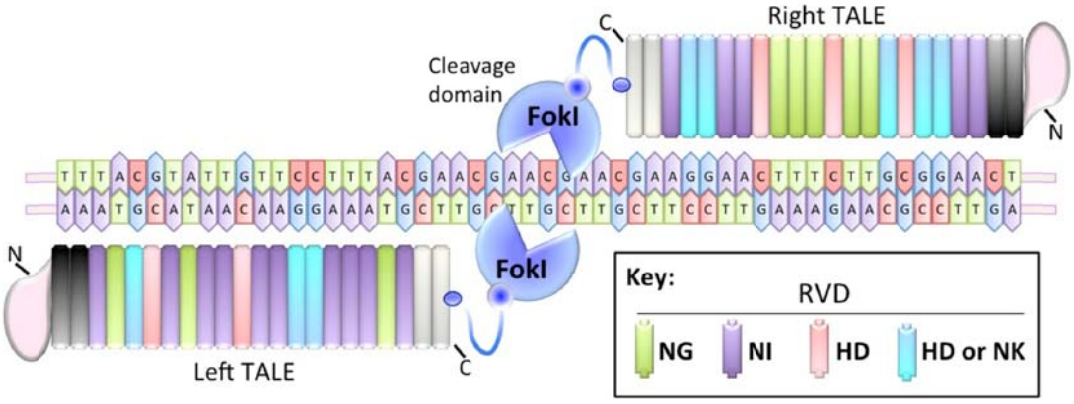


Fig. 1-4 Schematic diagram of TALENs components and genome-editing processes.

1.4.4 CRISPR/Cas9

1.4.4.1 The research history of CRISPR/Cas system

The discovery of CRISPR/Cas began in 1987 when Nakata and his colleagues [100] identified a cluster of 29 bp repeats downstream of the IAP gene in *E. coli*. These DNA sequences exhibited a unique cluster duplication, which was subsequently found in about 40% of bacterial species and 90% of archaea, such as *Mycobacterium spp*, [101, 102], *Salmonella entreica*, and *Shigella dysenteriae* [77, 103]. These repeat sequences were later referred to as non-repeat short sequence separation of spacer

DNA. In 2002, Jansen and Mojica called coined the term CRISPR for these short repeats. CRISPR is located in the prokaryote genome or some plasmids. The number of CRISPR loci and the number of repeat sequences within CRISPR varies among species [104]. Additionally, four genes encoding CRISPR-associated (Cas) proteins were discovered [105]. CRISPR/Cas system is speculated to play an important role in biological processes. Moreover, some of the spacer sequences are homologous with phage DNA sequences. These spacer sequences are indeed from phage DNA. After the bacteria are infected, the CRISPR system selects about 20 bp DNA fragments near the potential PAM (prospacer adjacent motif) sequence of phage genomic DNA and inserts them into the bacteria or archaea to extend the CRISPR expression box. Upon re-infection, the presence of these spacer sequences in bacteria can prevent the invasion of foreign plasmids or phages. Two research groups [106, 107] combined CRISPR with immune system to prevent the invasion of exogenous DNA and predicted that CRISPR might play an important role in immune defense through a mechanism similar to that of eukaryotic RNA interference. Subsequently, Barrangou et al. inserted the phage spacer into *S. thermophilus* in 2007, which conferred resistance to phage infection. This confirmed that CRISPR confers the bacteria with immunity against phage invasion [108]. In 2010, Garneau et al. demonstrated that the spacer sequence in *Streptococcus thermophilus* can guide Cas9 in *Cas* gene cluster to cut DNA [109]. These important findings enabled the scientists to study the mechanism of CRISPR/Cas9 system. Subsequently, there was a rapid development of the CRISPR/Cas9 system.

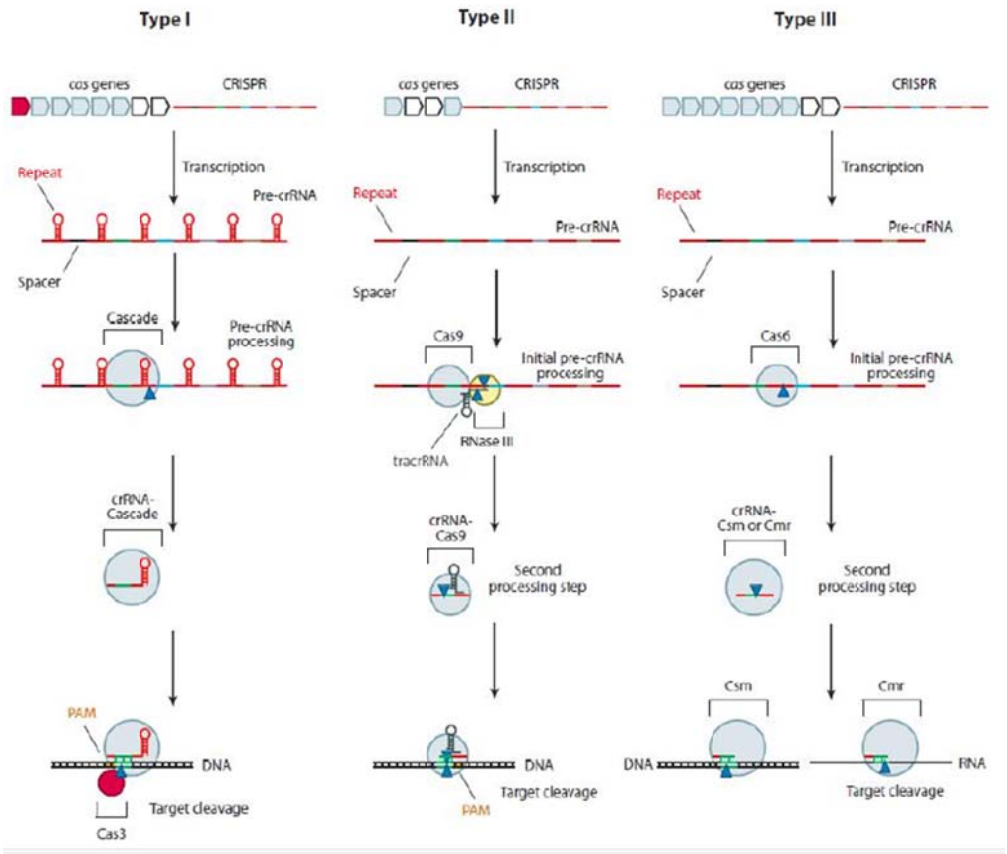


Fig. 1-4 Schematic overview of the Type I, II, and III CRISPR expression and interference stages [110]

1.4.4.2 The structure of CRISPR/Cas system

CRISPR/Cas system comprises the CRISPR sequence and *Cas* gene. The CRISPR sequence comprises the leader, repeat, and spacer. Most CRISPR sequences have a 5′ terminal with a length of 300–500 bp and an adenine-rich precursor sequence. The precursor sequence has a transcription start site, which can initiate the transcription of CRISPR sequence [111, 112]. However, the sequence lacks ORFs and thus cannot encode proteins. The precursor sequence is not strongly conserved. The sequence is relatively well-conserved within the same species with about 80% of the sequences being identical. The variability is very high in different species [105]. The repeat sequence is a 24–48 bp forward repeat sequence with 5–7 bp of palindrome sequences of GTTT/G and GAAAC at both ends. The mature CRISPR RNA (crRNA) formed after transcription has a stable neck ring structure [113] and forms a complex with Cas protein [114]. The repeat sequences are not continuous but are separated by

interval sequences of 26–72 bp. The bacteria with only one repeat can still have spacer sequences [115]. The number of interval sequences in different CRISPR sequences also varies. There are 587 interval sequences in the CRISPR sequence of a slime bacteria, which is the most number of known interval sequences. These sequences are not the genome sequences of bacteria and are from phage or plasmid DNA sequences. These sequences confer the bacteria with resistance against phages.

Cas gene, another important component of the CRISPR/Cas system, is located near the CRISPR sequence. The *Cas* gene is usually a group of conservative protein-coding genes in the upstream region, including nuclease, helicase, polymerase and RNA binding domains, which combine with crRNA to form ribonucleoprotein complex to specifically degrade foreign DNA through binding sites [116]. Generally, there is a corresponding Cas protein gene near the active CRISPR sequence. However, if there are multiple CRISPR sequences in the same bacteria, the CRISPR sequences without *Cas* gene in some adjacent regions can be transcribed and combined with the *Cas* gene at other positions in the genome to form a complex.

1.4.4.4 Type of CRISPR/Cas system

The bioinformatics analysis of advanced sequencing data revealed that the *Cas* gene exhibits high diversity and 45 Cas proteins have been reported successively [117]. The CRISPR/Cas system is often found in bacteria and archaea inhabiting extreme environmental conditions. The functioning of some CRISPR/Cas systems is conserved in all prokaryotes. However, several Cas protein families reflect a different evolutionary pattern. The continuous coevolution between viruses and their hosts has led to the emergence of anti-CRISPR in viruses, which may explain the high diversity of CRISPR/Cas. The CRISPR/Cas system is divided into three types based on the sequence and structure of Cas protein (type I, type II and type III) and at least 10 subtypes (Fig. 1-4).

Type I CRISPR/Cas comprises 6 different subtypes (I-A to I-F) in bacteria and archaea. Cas3 is the essential and major conserved marker protein in the interference reaction. Cas3 contains an HD phosphohydrolase region and a DExH-like helicase

region [118]. These two regions are independently encoded by two unrelated genes. Depending on the presence of ATP and Mg^{2+}, dsDNA is desorbed in the helicase region and ssDNA is cleaved in the HD region [119]. Cas3 interacts with different cascades (CRISPR-associated complex for anti-viral defense, cascade) and transports crRNA. The crRNA and cascade complex can recognize the target DNA sequence. The recruited Cas3 degrades the target virus DNA molecule by forming negative super helix DNA.

Type II CRISPR/Cas is the smallest Cas gene with unique characteristics and is only found in the bacterial genome. In such systems, the multifunctional protein Cas9 is involved in both crRNA maturation and subsequent interference response [120]. The process of crRNA maturation depends on the trans-activated crRNA (tracrRNA), which contains 25 nucleotides and crRNA repeats matched near the CRISPR site [121]. Cas9 can promote the pairing of tracrRNA and pre-crRNA base to form RNA double strand. RNase III cuts the double strand RNA to produce mature crRNA [122]. The cleaving of dsDNA requires crRNA, tracrRNA, and Cas9. The McrA/HNH nuclease region of Cas9 cuts the complementary DNA strand with crRNA. The RNase H folding region cuts the non-complementary DNA strand in the presence of Mg^{2+}. The DNA is cleaved precisely at a site that is 3 nucleotides upstream of PAM, while the non-complementary DNA strand is cleaved at 3–8 bp upstream of PAM sequence and produces a blunt end [109].

Two known CRISPR/Cas systems of type III (III-A and III-B) are mainly present in the genome of archaea. Type III system encodes CRISPR-specific ribonuclease-Cas6 protein and Cas10 protein subtypes, which are likely to participate in targeted interference. Cas10 encodes an area of targeted degradation of HD nuclease. The type III-A system of *Staphylococcus epidermidis* contains five Csm proteins that can target DNA [123]. This type of target DNA system does not need a specific PAM sequence but cannot target the sequence complementary to 8 nucleotides of crRNA [124]. In the type III-B system of *Pyrococcus furiosus*, Cas6 is not an essential part of the interference complex after the crRNA ripening process.

However, the 8 nucleotides of 5′ repeat sequence label provide the nuclear protein interference complex with the fixed assembly of 6 proteins (Cmr1–Cmr6). The Cmr complex containing 7 proteins (Cmr1–Cmr7) in *Sulfolobus solfataricus* exhibits endonuclease activity at the UA dinucleotide of the invading RNA. For both Cmr complexes, PAM sequences are not required for the target RNA. Unlike other subtypes, these two types of interference complexes specifically target RNA rather than DNA [125]. However, Cmr protein is reported to target the plasmid DNA in vivo independent of the PAM sequence.

1.4.4.5 CRISPR/Cas9 system as a genome editing tool

There is great potential to use different CRISPR/Cas systems as genomic manipulation tools. It is necessary to study the activity of different Cas proteins and use them for the development of different editing tools. For example, Cas6f, previously known as Csy4, is a pre-crRNA process enzyme that predicts gene expression. In recent years, studies on the Cas protein interference complex have revealed that Cas proteins play an important role in the development of new genome editing tools. They can be used to target specific DNA or RNA. One of the most important Cas proteins is Cas9, a large type II protein. During the early stage, natural Cas9-mediated genome editing is realized in two steps. Cas9 induces DSB in the genomic DNA target site through a 20-nucleotide guiding sequence in crRNA. Next, the DSB is repaired through NHEJ or HDR. The natural Cas9 system requires the following three basic parts: Cas9 nuclease, tracrRNA, and designable crRNA. The type II CRISPR/Cas system is further simplified with the system requiring only two parts: Cas9 nuclease and designable gRNA. Studies on Cas9 interference have revealed that the fusion product of crRNA and racrRNA has similar efficiency with the crRNA: tracrRNA double strand processed by RNase III [122]. Therefore, the design method of Cas9 and sgRNA is similar to the fused crRNA/tracrRNA sequence. The ribonucleoprotein formed is called RNA guided endonucleases (RGENs). RGENs can target a single gene or even multiple genes and edit the target sequence efficiently and specifically. The specificity of targeting is determined by the sgRNA sequence.

Different single Cas9 proteins of sgRNA can be targeted repeatedly without the time-consuming reassembly like protein-directed artificial nuclease. Several sgRNAs can target edit five genes at the same time in a single reaction. This method of genome editing opens a wider choice for genome editing of different species and different kinds of cells.

The Ruvc or HNH domain of mutant Cas9, and the gRNA mediated Cas9 can generate the gap at the target site. This mutant complex, gRNA mediated Cas9, can produce DSB- and NHEJ-mediated mutations at specific target sites when used in a pair. This double notch design can target the complementary chain of a target site to initiate HDR, which is more efficient and faster than natural Cas9-mediated HDR and single Cas9-mediated HDR.

1.5 Research purpose

It is anticipated that transgenic avian species will be used as living bioreactors for the production of biopharmaceutical proteins. Precise tissue-specific expression of exogenous genes is a major challenge for the development of avian bioreactors. No robust vector is currently available for an efficient and specific expression. In recent years, genome editing techniques, such as the CRISPR/Cas9 system have emerged as efficient and user-friendly genetic modification tools. To apply the CRISPR/Cas9 system for the development of transgenic chickens, guide RNA (gRNAs) sequences of the CRISPR/Cas9 system for the ovalbumin (*OVA*) locus were evaluated for the oviduct-specific expression of exogenous genes. An *EGFP* gene expression cassette was introduced into the *OVA* locus of chicken DF-1 and embryonic fibroblasts using the CRISPR/Cas9 system via homology-independent targeted integration. For the knock-in cells, *EGFP* expression was successfully induced by activation of the endogenous *OVA* promoter using the dCas9-VPR transactivation system. The combination of gRNAs designed around the *OVA* TATA box was important to induce endogenous *OVA* gene expression with high efficiency. These methods provide a useful tool for studies on the generation of transgenic chicken bioreactors and the

activation of tissue-specific promoters.

1.6 Thesis components

Chapter 1 introduces general information on the recombination protein expression system. Additionally, the difficulty in the generation of transgenic chicken is introduced. Furthermore, the mechanisms of gene editing methods are also discussed.

In Chapter 2, the techniques and mechanisms related to the study in this thesis are reviewed.

In Chapter 3, the selection of high efficiency gRNAs for the ovalbumin and lysozyme genes in the CEF cells is discussed. The HITI pathway was used to evaluate the knock-in efficiency of an exogenous gene.

In Chapter 4, 5 gRNA sequences around the ovalbumin gene promoter (TATA box) were co-cultured into CEF and DF-1 cells to initiate ovalbumin protein expression to check dCas9 transactivation system efficiency.

In Chapter 5, An *EGFP* gene expression cassette was introduced into the *OVA* locus of chicken DF-1 using the CRISPR/Cas9 system via homology-independent targeted integration to confirm the endogenous *OVA* gene promoter able to initiate exogenous gene expression with high efficiency

In Chapter 6, the contents of this study are summarized.

Chapter 2

Background

Because of the different of reproduction mechanism between mammal and poultry, transgenic technologies in mammal manipulation don't apply for transgenic avian. Depending on the chicken reproduction characteristics, many methods are used to prepare transgenic chicken, such as sperm-mediated injection, Stem cell culture, subgerminal cavity microinjection and PGCs-mediated injection. Even though all the methods have successful reports, but the efficiency still is unsatisfactory.

Beside gene delivery, the biggest problem of the research of transgenic chicken is specific expression of exogenous gene. To achieve specific expression, people formerly studied the structure of chicken ovalbumin genome 5-terminal regulatory element. But so far, the high specificity and efficiency of expression has not yet been achieved. Recently, the genome editing technologies are applied for transgenic research, utilizing the high practicability of those techniques is a new strategy for transgenic avian bioreactor.

2.1 DNA double strand breaks repair mechanism

Double strand breaks (DSBs) seriously threaten the integrity of the genome. There are two main mechanisms to repair DSBs: non homologous end joining (NHEJ) and homologous directed repair (HDR). Gene editing technology cut genome DNA by nuclease to form double strand breaks, and then insert, delete or replace bases through different repair mechanisms.

2.1.1 Homologous recombination mediated repair

Homologous recombination (HDR) is caused by the binding of MRN complex to DSBs. When DNA double strand breaks, the ends of DSBs are excised by the endonuclease (CTIP), resulting in the long 3 'single stranded DNA (ssDNA)

fragments being wrapped by the replication protein A (RPA). In the key step of HDR, Rad51 replaces RPA to bind with ssDNA to form presynaptic filaments of nucleoprotein and promote the search for homologous donors. Rad51 after the formation of a heteroduplex DNA structure decomposition, along with DNA synthesis and the final linking steps [126].

Gene editing induced by Cas9 system with HDR repairmen require three parts, including Cas9 or its variants protein (used to generate double strand breaks), guiding RNA which to guide Cas9 protein to genome target region) and template contain homologous sequence depends on editing sequence locus used to mediated HDR repaire pathway. Using these three components, researchers can bring deletion or insertion to knock-out genes, and design nucleotide changes to correct disease phenotypes. In addition, HDR editors can disrupt gene expression by inserting genetic markers into genes, such as GFP or puromycin screening expression box. In addition, safe sites can be inserted with complete genes cassette, including promoters, CDS and polyA signals. The insertion of complete genes (such as drug resistance genes or fluorescence genes) can be used to screen target cells, and is also conducive to the study of protein structure or function (Fig. 2-1).

2.1.2 Nonhomologous end joining mediated repair

The non-homologous end joining (NHEJ) signaling pathway has been studied in simple and higher eukaryotes. Although the research methods are different, it is found that the molecular mechanism is almost the same, and the process can be roughly divided into three steps: first, Ku complex recognizes and binds to the end of DNA double strand break to protect DNA from degradation by nuclease [127]. Then DNA-PKcs (protein kinase catalytic subunit) protein was collected to form DNA-PK whole enzyme trimer [128]. DNA-PKcs can self-phosphorylate and activate DNA ligase VI and XRCC4 protein when it binds to DSB terminal, and at the same time, Artemis protein can be recruited. Artemis protein can remove 5 'and 3' protruding terminal and hairpin structure of any length [129]. Finally, DNA ligase VI and

XRCC4 complex jointly complete DNA connection. In the process of NHEJ repair, many proteins are involved in DSB recognition and connection. At present, it is known that there are six key genes encoding proteins directly involved in this pathway, namely *XRCC4*, *XRCC5*, *XRCC6*, *LIG4* and *DCLREIC*. *XRCC5* and *XRCC6* encode 86 kDa Ku80 protein and 73 kDa Ku70 protein, respectively, which can form heterodimer protein of circular structure, referred to as Ku complex. Ku complex was originally found as an autoantigen in patients with multiple myositis syndrome [130], which is free to bind to the end of DSB without selectivity. Ku complex as a sensor of DSB exists in the nucleus. It was found that when *XRCC6* or *XRCC5* gene was knocked out, the mice showed growth retardation, radiation sensitivity, and decreased recombination efficiency of V(D)J, which indicated that *XRCC6* and *XRCC5* played an important role in the NHEJ signaling pathway [131, 132]. *PRKDC* encoded DNA-PKcs is a serine and threonine protein kinase, belonging to phosphatidylinositol 3-kinase protein family, which is only found in vertebrates [133]. DNA-PK holoenzyme composed of DNA PKcs and Ku complex is the key enzyme in NHEJ signal pathway. DNA PKCs can not only self-phosphorylate to change its conformation, but also promote the binding of other proteins with DSB, and also phosphorylate a variety of proteins involved in the process of apoptosis [134]. It was found that DNA PKcs gene knockout resulted in mouse embryo death [135]. *XRCC4* encodes a nuclear phospholipid protein composed of 334 amino acids, 75% are homologous between human and mouse. *XRCC4* can damage DNA binding of *LIG4* and enhance the activity of *LIG4* [136]. Recently, it was found that *XRCC4* gene mutation can cause mouse embryo death [137]. *LIG4* is composed of two parts. It is an ATP dependent DNA ligase. It can not only connect the complementary gaps on single or double chains, but also connect the non-complementary ends with the participation of *XRCC4* and Ku complex [138]. *DCLREIC* encodes Artemis protein, which is activated by DNA PKcs phosphorylation and participates in the pre connection of NDA terminal. Once the end of DNA break is processed, it will leave the Ku complex with DNA PKcs and provide the binding site for other proteins. The

absence of Artemis also caused RS-SCID in mice [139].

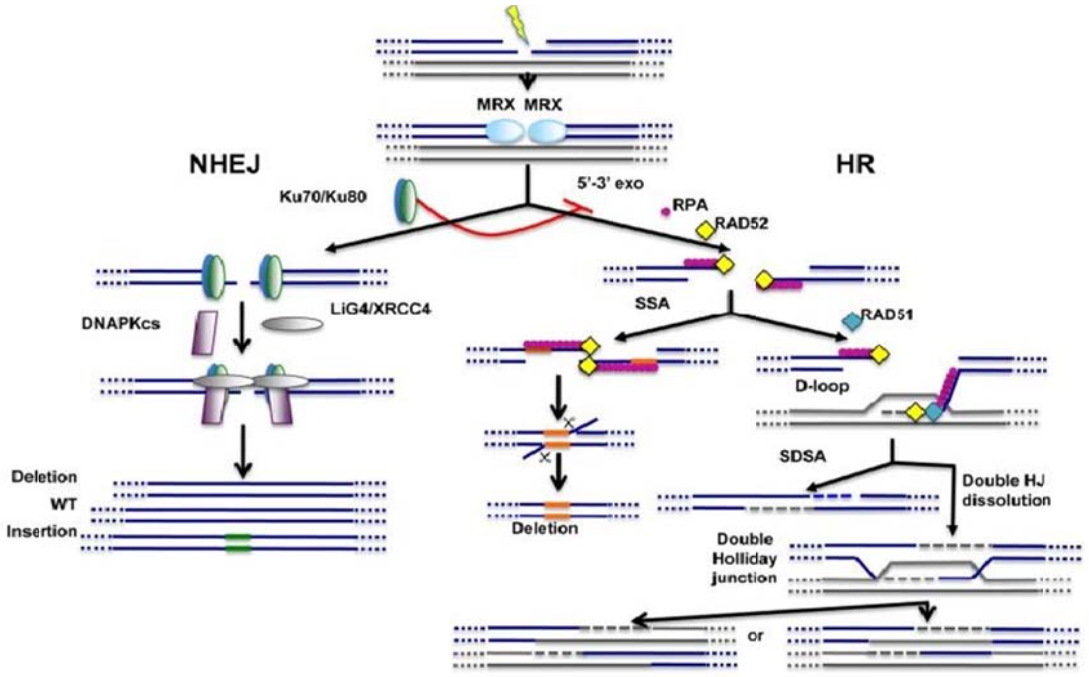


Fig. 2-1 Main pathways of DNA repair [140]

2.1.3 Microhomology mediated end joining mediated repair

In addition to the classic NHEJ (cNHEJ) that depends on Ku70/80, there are other NHEJ pathways in cells that do not rely on Ku70/80, called alternative NHEJ (aNHEJ). Microhomology mediated end joining (MMEJ) is a relatively clear type of a-NHEJ [141, 142]. Compared with cNHEJ, MMEJ has three characteristics. First, it cannot rely on the classical NHEJ repair protein, for example, the repair process cannot rely on Ku, for example, ligase III/I replaces ligase IV to perform the connection. Second, the repair results contain more errors, such as deletion of long segments, and more chromosomal aberrations [143]. Third, it depends on the homology of short sequences, which can be produced by nuclease cleavage or pol μ cross injury repair, and the microhomology is also the structural basis of the classical NHEJ protein. In addition, MRN complex and CTIP (CtBP interacting protein) of HR pathway can promote MMEJ [142]. MMEJ was originally considered as a complementary form of cNHEJ and only in the absence of cNHEJ.

It was found that MMEJ can coexist with cNHEJ in a low proportion [144]. There

are still disputes about whether MMEJ is a special form of NHEJ or a new way independent of cNHEJ.

2.2 Detection of knock-out mutation

2.2.1 T7endonuclease I assay

With the development of gene editing technology, rapid, efficient and economic methods to detect mutations are particularly important. Traditional gene sequencing technology is expensive and time-consuming, often cannot meet the needs of high-throughput mutation detection. Recently, the high-resolution melting curve has been proved to be able to detect mutations, but this method needs a special quantitative method for small PCR products. Another effective method to detect mutations is to hybridize the mutated DNA sequence with the non-mutated DNA sequence to form heteroduplex. Specific endonuclease can detect and cut the mismatched DNA sequence. T7endonuclease I (T7E1) is the specific endonuclease for gene mutation detection. T7E1 endonuclease can recognize and cut incomplete matched DNA, cross structured DNA, Holliday structured or cross DNA, heteroduplex DNA or double stranded DNA at a slower speed.

There is also a key control factor in the detection of mutated genes by endonuclease mismatch shearing method, that is, the process of PCR amplification of template DNA. If the fidelity of DNA polymerase is limited, then the amplified DNA fragments may introduce mutations, especially for the cycle of PCR amplification end, which will lead to the formation of recombinant PCR products, thus leading to the following experimental results are affected. The solution to this problem is to use high fidelity DNA polymerase to amplify DNA fragments, to ensure the specificity of the amplified fragments, and to reduce the irrelevant background caused by DNA polymerase chimerism error.

2.2.2 Single strand annealing recombination assay

All reporting systems are based on special reporter genes, such as antibiotic

resistance genes, fluorescent protein coding genes and luciferase genes. The target sequence of specific nuclease is inserted into the open reading frame (ORF) of the report gene, so that it cannot correctly express the corresponding functional protein. Under the action of specific nuclease, the target sequence inserted in the report gene is targeted to be cut to form DSBs, and according to the design of researchers, the internal repair mechanisms such as NHEJ, single strand annealing (SSA) or HR of cells are activated to realize the report. Repair of ORF gene. The report system is easy to observe and detect. SSA is an important way of DNA recombination. It is activated when there is a certain length of positive repeats on both sides of DSBs. It can repair DSBs by homologous recombination and delete a positive repeats. In theory, after the introduction of DSBs, the repair efficiency of the reporter gene based on SSA repair mechanism can reach 100%.

In order to facilitate observation and detection, the gene encoding fluorescent protein is generally used as the reporter gene. Ramakrishna fused and expressed the red fluorescent protein mRFP and green fluorescent protein *EGFP* for the first time, inserted ZFNs target sequence between their coding gene sequences, and constructed the mRFP *EGFP* double fluorescent report vector. Among them, mRFP was correctly expressed as a marker gene to measure the transfection efficiency, while *EGFP*, as a reporter gene, had a frameshift mutation and could not be normally expressed. When ZFNs cut the target sequence and introduce DSBs, it will activate the SSA repair mechanism in cells, rearrange the reading frame of *EGFP*, and express the functional green fluorescent protein. The ratio of *EGFP* positive cells was used to measure the working efficiency of ZFNs, and the double positive cells of mRFP and *EGFP* were further separated by flow cytometry to enrich the positive cells modified by genome [145].

2.3 CRISPR/dCas9 system in gene expression regulation

CRISPR/Cas9 nuclease is widely used as a gene editing tool due to its activity. In addition, the cleavage activity of CRISPR/Cas9 nuclease can be removed by mutation,

and it can be turned into a tool to collect protein and RNA to specific sites, thus becoming a powerful sequence specific tool to regulate gene expression [146-149]. At the same time, dCas9 remains the characteristic of binding to specific sites through sgRNA, so it becomes a tool for regulating gene expression.

Compared with the existing gene expression regulation tools, such as RNA interference (RNAi), tale -, ZF -, CRISPR/dCas9 has the following advantages: simple and convenient, easy to design and apply, high sequence specificity, direct regulation of gene expression at the transcription level, and the coding region that also act on the non-coding region [146]. CRISPR/dCas9 system can achieve very efficient activation or inhibition in a very specific small area, and is very sensitive to the mismatch between sgRNA and DNA （Fig. 2-2）[150].

CRISPR/dCas9 gene expression system showed little off-target in different systems in the whole genome screening experiment [150, 151]. At present, studies have been performed to apply CRISPR/dCas9 gene regulatory system to different organisms to regulate their gene expression. Like flies, plant [152] can also be used to keep HIV-1 permanently silent [153].

2.3.1 CRISPR / dCas9 system suppresses gene expression

Due to the lack of RNAi mechanism in bacteria, this simple method of regulating bacterial gene expression cannot play a role. The sequence specific gene inhibition based on Cas9 was first found in *E.coli*, known as CRISPRi. dCas9-sgRNA complex can recognize specific sites, prevent gene transcription by blocking the extension of RNA polymerase (Pol), or occupy the transcription starting site, and destroy the formation of transcription complex. CRISPRi can effectively inhibit gene expression in bacteria, but it has little miss target effect. Using multiple sgRNA, CRISPRi can inhibit multiple genes simultaneously in bacteria. Different from CRISPR/Cas9 as a gene editing tool, CRISPRi is reversible for the gene inhibition [149]. In bacteria, CRISPRi can be used as an efficient tool to inhibit gene expression, indicating that this technology has the potential to be applied to more species.

The first experiment that CRISPRi was applied to mammalian cells was to inhibit the expression of *GFP* by using the sgRNA of dCas9 binding target to *GFP* in 293T cells which stably expressing *GFP* [149]. It was found that the effect of inhibiting gene was not obvious when only dCas9 was used in mammalian cells. In order to enhance the gene inhibition effect, scientists fused the inhibition factor KRAB or SID4X with the carboxyl end of dCas9 to form dCas9-KRAB, dCas9-SID4X. These fusion proteins can effectively inhibit the expression of endogenous genes in mammalian cells. Some researchers also fused the amino terminus of KRAB and dCas9, which can also inhibit the expression of endogenous genes in mammals [150]. The inhibition effect of dCas9 or dCas9 fusion protein on gene is sgRNA site dependent, which is likely to be the three-dimensional structure of chromosome and the distribution of inhibition factors, making some sites have obvious inhibition effect, while others have not. In yeast, the inhibitor Mxil is used to inhibit the expression of endogenous genes. However, the application of CRISPRi in eukaryotes needs further optimization, because the current system can only achieve the expression of suppressor genes at some sites.

2.3.2 CRISPR / dCas9 system activates gene expression

By fusing the activator VP64 or p65AD with dCa9, the resulting complex can activate endogenous genes in mammalian cells, a system known as CRISPRa. However, the researchers found that to activate endogenous genes such as *IL1RN*, *ASCL1*, *NANOG*, *MYOD1*, *VEGFA* and *NTF3* with compounds like dCas9-VP64 or dCas9-p65AD, multiple sgRNAs are required to work together to observe the obvious effect [154]. Through protein engineering, the researchers improved the activation effect of CRISPRa in endogenous genes. For example, scientists found that the combination of VP64 with the carboxyl and amino ends of dCas9 can achieve better activation effect, and the fusion expression of multiple VP16 and dCas9 to form dCas9-VP160 can play a better activation role. However, multiple sgRNAs are still needed for *IL1RN*, *Oct4* and *Sox2* to be activated. In large-scale gene screening with

CRISPRa, only one effective sgRNA can be used for each gene. For a single site, more than one sgRNA is needed to activate. The experiment shows that recruiting more than one activator in one site can enhance the activator. Therefore, some researchers fused more than 10 repetitive polypeptides GCN4 and the carboxyl end of dCas9 in the SunTag system, to fused scFv, GFP protein and VP64 to form two complexes of dCas9-10XGCN4 and scFv-sfGFP-VP64, to realize the effective activation of endogenous genes through the SunTag system. Using this system, the researchers can effectively activate the CXCR4 endogenous gene with only one sgRNA, and observe the obvious cell migration [155]. Another group tried to find a more efficient way to activate genes by screening different activators. The research team finally developed VPR system, which fused VP64, p65AD and Rta to the carboxyl end of dCas9 to form dCas9-VPR complex. Compared with dCas9-VP64, dCas9-VPR can activate genes in coding and non-coding regions more efficiently. This system can activate endogenous genes in drosophila, yeast and mouse cells [156].

Through sgRNA engineering, scientists also try to improve the activation efficiency of CRISPRa on endogenous genes. In the sgRNA structure, insert a RNA aptamer sequence which can be recognized by RNA binding proteins (RBPs), and fuse the activators VP64 and p65AD with RNA binding proteins to form a new activation system. The researchers used this system to activate genes through multiple sgRNAs in endogenous genes [157]. Using this technology, scientists developed SAM activation system, inserted MS2 sequence into the skeleton of sgRNA, and fused the recognition protein MCP of MS2 sequence with p65AD and HSF1. The combination of SAM system and dCas9-vp64 can activate endogenous genes more effectively.

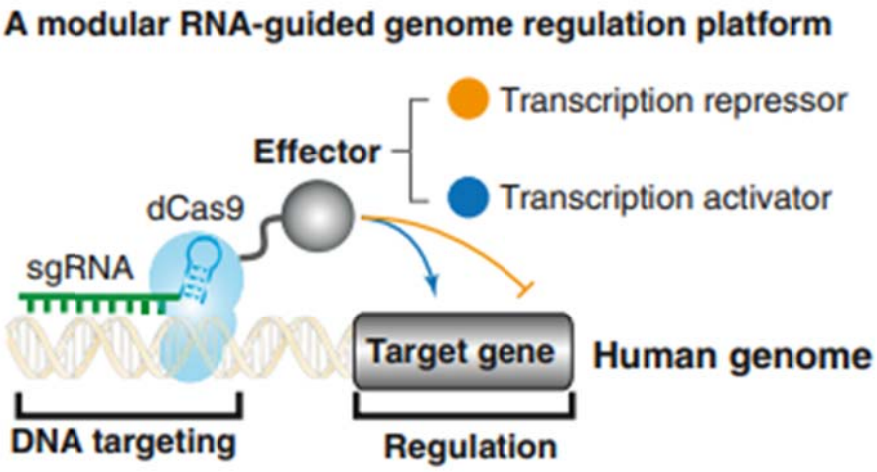


Fig. 2-2 dCas9 as a modular system for attachment of transcriptional regulators [148]

Conclusion

The function analysis of 5' and 3' terminal regions of the genomic locus of ovalbumin (*OVA*) gene revealed the location of regulatory elements such as promoter and enhancer, and the information is essential for constructing transgene expression and/or knock-in vectors for oviduct-specific transgene expression. The regulatory elements of the *OVA* 5'-terminal region are crucial for accomplishing specific expression. There are four DNase I hypersensitive sites in the regulatory region [64], in which the steroid-dependent regulatory element (SERE) and the negative regulatory element (NRE) which locates downstream of SERE associate with expression control. In the report, SERE locates from −892bp to −780bp, initiates gene expression of oviduct epithelial cell via binding with estrogen and glucocorticoid cooperating with NRE between −308bp to -88bp [57, 59, 60]. Thus, *OVA* expression by oviduct cells is tightly regulated using the endogenous mechanism. Here we used CRISPR/Cas9 system and HITI approach for knocking-in an *EGFP* expression cassette with or without *CMV* promoter into the *OVA* locus of DF-1 and chicken embryonic fibroblast (CEF) cells, in which *EGFP* expression directly under *CMV* promoter or not. After gRNA sequences of dCas9-VPR transactivation system for activating endogenous *OVA* gene in DF-1 and CEF cells were designed, we confirmed *EGFP* expression of the knock-in cells.

Chapter 3

Selection of high efficiency gRNAs in *ovalbumin* and *lysozyme* genes and determine HITI pathway to knock-in exogenous gene efficiency

3.1 Introduction

At present, the widely used CRISPR Cas system is composed of s main parts: crRNA with guiding function, trans-activated crRNA and Cas nuclease protein. In 2012, the research team of Doudna and Charpentier combined crRNA with tracRNA, and designed a single RNA molecule as the guide RNA (gRNA) to guide the direction cutting of Cas protein, which is known as CRISPR/Cas9 system. The protein encoded by the gene has the activity of helicase, which contains the HNH active site responsible for cutting the complementary sequence with gRNA, and the Ruvc active site responsible for cutting another free DNA chain with the same gRNA sequence. Determined by the simple structure of CRISPR / Cas9 system, its specific recognition is only composed of 20 variable region gRNAs with target specificity, and NGG (protospace fajacent motif, PAM) sequences used to connect the reporter sequence and Cas9 protein.

In addition to 5' regulatory sequences, 3' regulatory sequences and introns of some genes have different effects on tissue-specific expression and level. Palmiter et al.[158] experiment with transgenic mouse model the results showed that introns, especially the first intron, were necessary for the efficient expression of transgenes. Research results showed that the expression vector constructed by the 5' regulatory region of 3.0kb *OVA* gene including intron 1 and the 3.0 kb 3' regulatory region starting from exon 7 could not only effectively drive the expression of lacZ reporter gene and HK1 and other target genes in chicken oviduct epithelial cells, but also have good tissue-specific and estrogen dependent expression [159], indicating that two

regulatory sequences include the major proximal regulatory elements for the tissue-specific expression of *OVA* gene.

For quite a long time, crispr-cas9 system can only use normal replication mechanism in mitotic cells (such as those in the skin or intestinal tract), but not in mitotic cells, which cannot insert exogenous DNA into the target site of genome. A new pathway of the NHEJ mediated targeted integration, or homology-independent targeted integration (HITI), could repair DSB with robust donor vector and knock in genes in dividing or non-dividing cells [160]. HITI shows high knock-in efficiency, error free repair and low off target effects in vitro and in vivo [161].

3.2 Experimental Purpose

Considering the importance of gRNA sequence in our research, selection of high cut efficiency and low off-target gRNA sequence is necessary. In the present study, we first designed a series gRNA sequence in ovalbumin and lysozyme gene to select high cut efficiency of gRNA sequence in CEF cells. Then we build HITI mediated donor vector to knock in an *EGFP* expression cassette with *CMV* promoter into the *OVA* locus of DF-1 cells, with *EGFP* expression directly under the control of the *CMV* promoter to determine HITI pathway efficiency.

3.3 Selection of high efficiency gRNAs in *ovalbumin* and *lysozyme* genes

3.3.1 Material and Methods

3.3.1.1 Plasmid construction

Plasmids of pX330-U6-Chimeric_BB-CBh-hSpCas9 (pX330) encoding human codon-optimized SpCas9 under the control of the constitutive CBh promoter and a cloning site of chimeric guide RNA expression was obtained from Addgene (plasmids 42230). Oligonucleotides for gRNA templates were chemically synthesized (Invitrogen, Carlsbad, CA, USA), phosphorylated, and annealed in line with standard protocols. The gRNAs for exons, introns, and the promoter region of the ovalbumin and lysozyme gene were designed using CRISPRdirect software (http://crispr.dbcls.jp)

[162]. For SSA assay vector structure, *EGFP* gene was amplified with two group primer which PCR product one was 1~600bp of *EGFP* and another was contain 121~720bp,. After purified two PCR products and introduced into pcDNA4/IRES-*EGFP* which deleted IRES-*EGFP* part, pcDNA4-EGxxFP was prepared. Depends on designed gRNA location, amplified ovalbumin and lysosome gene target site region and ligated into pcDNA4-EGxxFP, totally get pcDNA4-EGxxFP-*OVA*1/2, pcDNA4-EGxxFP-*OVA*3/4, pcDNA4-EGxxFP-lys1, pcDNA4-EGxxFP-lys2, pcDNA4-EGxxFP-lys3/4.

The sequences used for gRNA are summarized in Table 3-1

The primer pairs used for vector construction are summarized in Table 3-2, restriction enzyme sites was underline.

Table 3-1. Sequences used for gRNA

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')[a]
gRNA/*OVA*1[a]	CACCGccagatgccaatctcgtaat	AAACattacgagattggcatctggC
gRNA/*OVA*2	CACCGtgcactggtgaccactagag	AAACctctagtggtcaccagtgcaC
gRNA/*OVA*3	CACCGacccagagcactgtagtatc	AAACgatactacagtgctctgggtC
gRNA/*OVA*4	CACCGtactggtaactcatggatga	AAACtcatccatgagttaccagtaC
gRNA/lys1	CACCGtgatagctggaagtcgctag	AAACctagcgacttccagctatcaC
gRNA/lys2	CACCGcaggctacagaacgggtatc	AAACgatacccgttctgtagcctgC
gRNA/lys3	CACCgtagcgtcgcgctcgcaaag	AAACctttgcgagcgcgacgctac
gRNA/lys4	CACCGcgagcgcgacgctacccgct	AAACagcgggtagcgtcgcgctcgC
gRNA/*OVA*-Intron1	CACCGtgctgtacatagtaccatgc	AAACgcatggtactatgtacagcaG
gRNA/*OVA*-Intron2	CACCGtgctgtggctccattgagc	AAACgctcaatggagccacagcaC
gRNA/*OVA*-Intron3	CACCGcaagacagctagatgattc	AAACgaatcatctagctgtcttgC
gRNA/*OVA*-Intron4	CACCGacactactaaatacactata	AAACtatagtgtatttagtagtgtC
gRNA/*OVA*-Intron5	CACCGagcttgaacgcaaagcacgc	AAACgcgtgctttgcgttcaagctC
gRNA/*OVA*-Intron6	CACCGtctctctttttttttttttt	AAACaaaaaaaaaaaaagagagaC
gRNA/*OVA*-Intron7	CACCgcaggtatggccctagaagt	AAACacttctagggccatacctgc
gRNA/*OVA*-exon2	CACCGagctctagccatggtatacc	AAACggtataccatggctagagctC

gRNA/*OVA*-exon3	CACCGtcaatactgtctccgaatcc	AAACggattcggagacagtattgaC
gRNA/*OVA*-exon4	CACCGttttagagcaacttactggc	AAACgccagtaagttgctctaaaaC
gRNA/*OVA*-exon5	CACCGccagagagctcatcaattcc	AAACggaattgatgagctctctggC
gRNA/*OVA*-exon6	CACCgtttgagaatccacggagct	AAACagctccgtggattctcaaac
gRNA/*OVA*-exon7	CACCGcaacatgctcattgtcccac	AAACgtgggacaatgagcatgttgC
gRNA/*OVA*-exon8	CACCgccaaagaagagaacggcgt	AAACccttcctttcttaagccctgC

[a] Target gRNA sequence is in lowercase

Table 3-2. Primer sequences used for vector construction.

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')[a]
EGxxFP1	TTTAAGCTTGCCGCCACCATGGTGAGCAAGGGCGAG	AAGGATCCTCAGTGGTTGTCGGGCAGCAGCACGG
EGxxFP2	AAggatccgctagcctgcaggtcgacgaattcgatatcGGCAAGCTGACCCTGAAGTTCAT	AAACTCGAGTTACTTGTACAGCTCGTCCATGC
Intron1-1cut[c]	GGGTGCTGTACATAGTACCATGCAGGTTCGCGATGTACGGGCCAGA	GGCAGTGAGCGCAACGCAAT

[b] Restriction enzyme site sequence is in lowercase.

[c] gRNA/Intron1 sequence is underline.

3.3.1.2 Cell culture and transfection

The chicken embryonic fibroblasts (CEFs) derived from 12-day-stage embryos (Line-M; Nisseiken, Yamanashi, Japan) and HEK293 cells were cultured in highglucose D-MEM (SigmaeAldrich, St. Louis, MO, USA) supplemented with 10% (v/v) fetal bovine serum (BioWest, Nuaillé, France), 100 units/mL streptomycin sulfate (Fujifilm Wako Pure Chemical Industries, Osaka, Japan), and 100 mg/L penicillin G potassium (Fujifilm Wako) in tissue culture dishes (Thermo Fisher Scientific, Waltham, MA, USA). DF-1 cells and CEFs were cultured at 37°C, in a 5% (v/v) CO_2 incubator.

To evaluate editing efficiency in the *OVA* and *Lys* gene, CEFs were transfected with pX330/gRNA vectors using Nepa21 electroporator (Nepagene, Chiba, Japan), in accordance with the manufacturer's instructions. Briefly, 2 days before transfection, CEFs were seeded into 100-mm culture dishes (Thermo Fisher Scientific). CEFs at a

density of 1×10^6 cells in 90 mL of Opti-MEM (Invitrogen) were transfected with 8 μg of each pX330 plasmid mixed with 10 mL of Opti-MEM. After electroporation, the cells were seeded into six-well tissue culture plates (Thermo Fisher Scientific).

To evaluate editing efficiency in the *OVA* and *Lys* gene by SSA assay, HEK293 cells were transfected with each pX330/gRNA vector and related pcDNA4-EGxxFP plasmid using Lipofectamine 2000 (Invitrogen), in accordance with the manufacturer's instructions. Briefly, 2 days before transfection, HEK293 cells were seeded into 100-mm culture dishes (Thermo Fisher Scientific). HEK293 cells at a density of 1×10^6 cells in 90 mL of Opti-MEM (Invitrogen) were transfected with 4 μg of each pX330 and each pcDNA4-EGxxFP plasmid mixed with 10 μL of Opti-MEM. After 48h transfection, the cells were subjected to flow cytometry analysis (SH-800; Sony, Tokyo, Japan) to observe *EGFP* expression.

3.3.1.3 Genomic PCR analysis

Genomic DNA was extracted from cells using a commercially available kit (MagExtractor Genome; Toyobo). Genomic DNA (50 ng) was subjected to PCR using DNA polymerase (KOD Plus Neo; Toyobo) at 94°C for 1 min; followed by 38 cycles of amplification at 95°Cfor 10 s, 56–60°C for 30 s, and 72°C for 20 s; and final extension at 72°C for 2 min. When the Tm value of primers was higher than 62°C, two-step PCRwas performed (95°C for 10 s, 68°C for 30 s). The primers used are summarized in Table 3-3.

Table 3-3. Primer pairs used for genomic PCR.

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')[a]
P-gRNA/Intron1	GCCTGATTAACTTCTAGCCCTAC	GTTGCACACTACTTGCTATGAG
P-gRNA/Intron2	CTGGCTTCTGGGACAGTTTGCTACC	TACCTGAGCTTCAATACTGTCTCC
P-gRNA/Intron3	GATAAACTTCCAGGATTCGGAGAC	TAGAGCAACTTACTGGCAGG
P-gRNA/Intron4~5	CTTCACTTAGAGACATCCTCAACC	TATCCCTACCATCCCTTAACTCC
P-gRNA/Intron6~7	AATGTCCTTCAGCCAAGCTC	ACCCAGTGGGACAAATCTAC
P-gRNA/Exon3	ACTGTTCTGCTGTTTGCTCTAG	TGCCTACCTGCTGCAATAATC
P-gRNA/Exon6	TGGTACTGTTTGGGTTGAAGAC	ACACAGTGTCTGAAAGAGATGC

P-gRNA/Exon8	CTCATTCTCATTTCCTTGCAGC	TGACAGCATAGGAATGGTTGG

3.3.1.4 T7EI assay

To detect insertion/deletion (indel) mutation by T7 endonuclease I (T7EI) mismatch assay, cells transfected with or without pX330-series plasmids were harvested for genomic DNA extraction. Genomic PCR products as described above were applied on a 1% (w/v) agarose gel to check for single-band products and purified using a PCR purification kit (NucleoSpin Gel and PCR Clean-up; MachereyeNagel). To induce DNA heteroduplex formation, purified PCR products (200 ng) mixed with 2 mL of NEB buffer 2 (New England Biolabs, Ipswich, MA, USA) up to 19 mL in a PCR tube were subjected to a denaturing/reannealing reaction using the following conditions: 95°C for 2 min, ramping down from 95°C to 85°C at -2°C/s, 85°C–25°C at -0.1°C/s, and then holding at 4°C. Reaction samples supplemented with 1 mL of T7 endonuclease I (New England Biolabs) were incubated at 37°C for 1 h. To stop the reaction, 2 mL of 250 mM EDTA was added. Finally, T7EI digestion products were analyzed using conventional gel electrophoresis. The indel mutation rate was calculated in accordance with a previous report [163].

3.3.2 Results and Discussion

3.3.2.1 CRISPR design to select gRNA in *ovalbumin* and *lysozyme* genes

Advances in genome editing technology have made it easier to generate transgenic animals. Unlike direct injection into single cell-stage embryos, it is necessary to introduce plasmid vectors of the CRISPR/Cas9 system with high efficiency into chicken embryos at the blastodermal stage (just after oviposition) by subgerminal microinjection because a chicken blastoderm contains more than $5\sim6\times10^4$ cells and drug selection cannot be used to eliminate non-knock-in cells. With the help of gRNA design software, we could quickly obtain high-score gRNA sequences and predict cleavage efficiency. We first designed each 4 gRNA sequences of CRISPR/Cas9 around the *OVA* and *Lys* 3' UTR region, in a strategy of co-expresses with endogenous *OVA* by IRES or 2A peptides. The location of gRNA

designed for transgene knock-in into the ovalbumin and lysozyme gene locus was shown in Fig. 3-1, and the sequences used for gRNA are summarized in Table 3-1. Depend on gRNA design software, total 8 gRNA sequences in *OVA* and *Lys* gene were high score of cleavage efficiency and with low off-target.

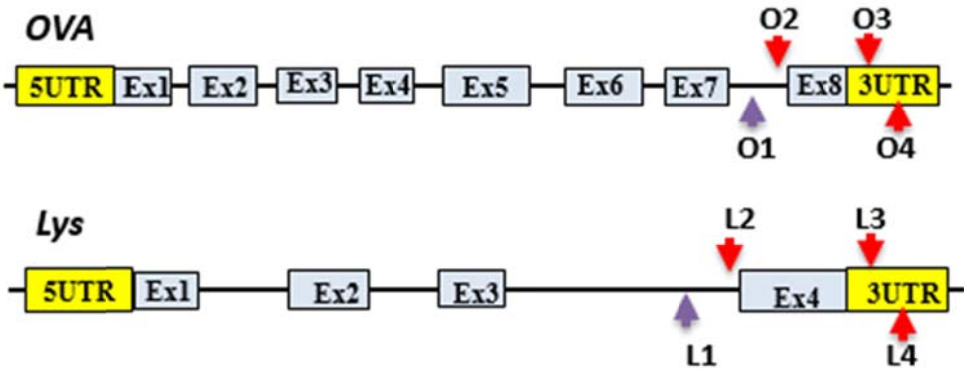


Fig. 3-1 The location of gRNA designed for transgene knock-in into the ovalbumin and lysozyme gene locus

3.3.2.2 SSA assay to test gRNA KO efficiency

For test gRNA sequences bring DSB efficiency, we using SSA assay to transform cut efficiency to *EGFP* expression efficiency which observe by microscope and *EGFP* expression cells are counted by FACS. After 48 h co-transfected SSA assay report plasmid and pX330 vector which express gRNA sequence into HEK293 cells. The schematic drawing of SSA assay to detected gRNA KO efficiency was showed in Fig. 3-2., the *EGFP* expression result were showed in Fig. 3-3, include *EGFP* expression cells were counted by FACS, the data was showed in Table 3-4. Depend on result, gRNA/*OVA* 2 and gRNA/Lys1 in SSA assay have most high cut efficiency, which are 80.92% and 69.40%, respectively.

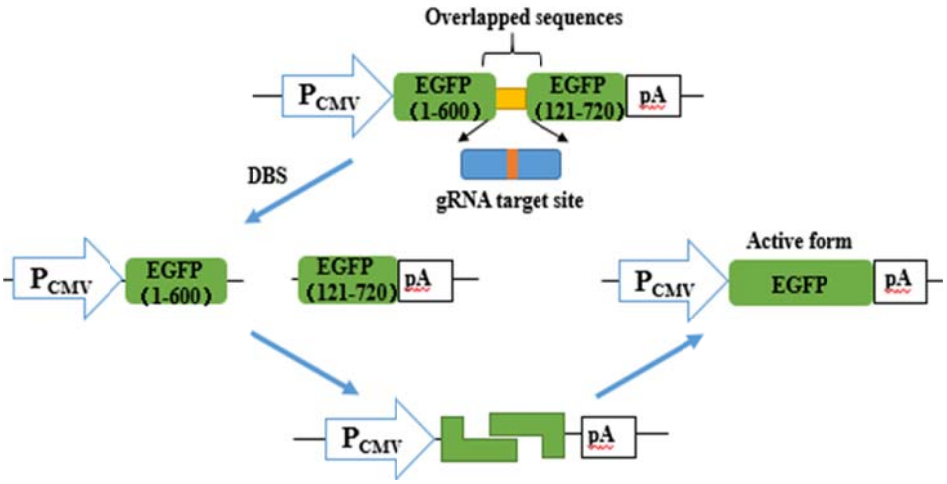
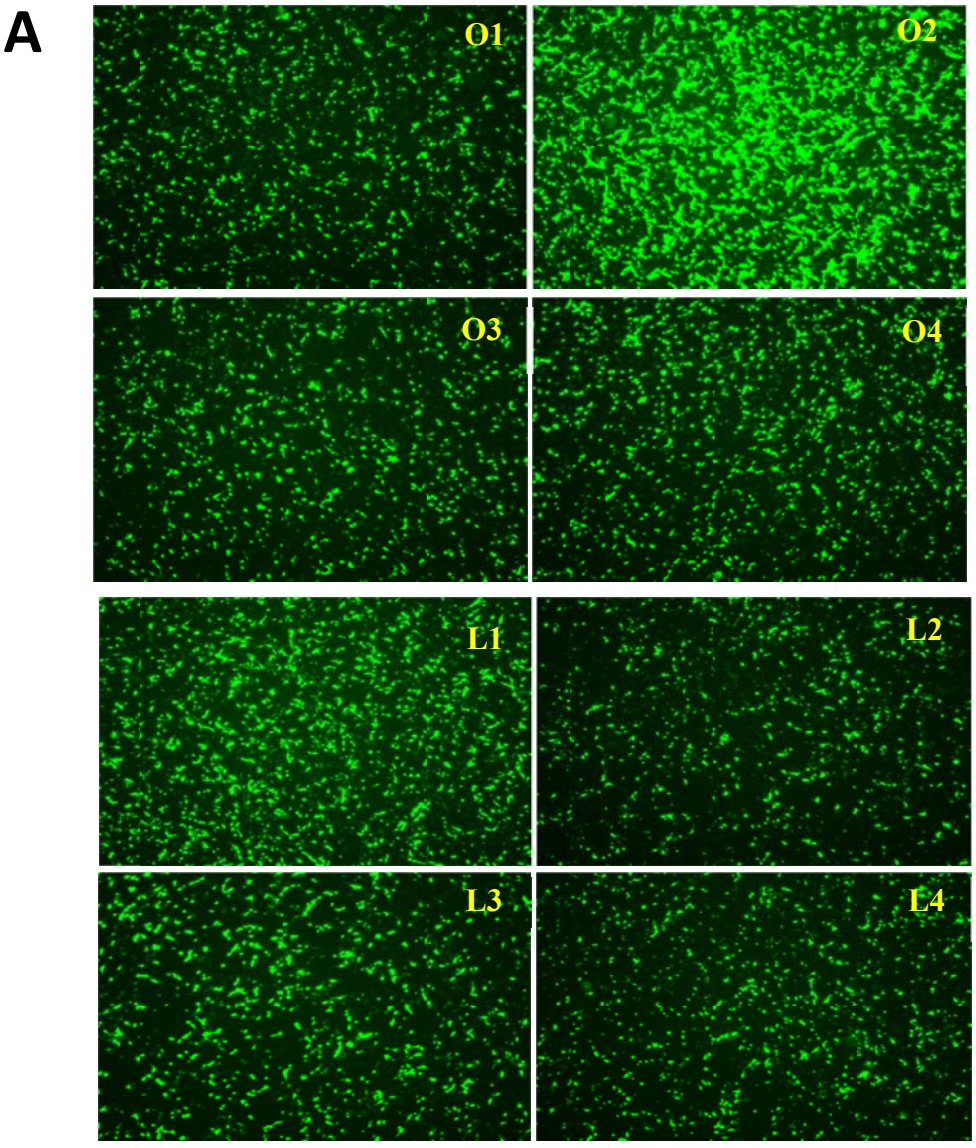


Fig. 3-2 The schematic drawing of SSA assay to detected gRNA KO efficiency



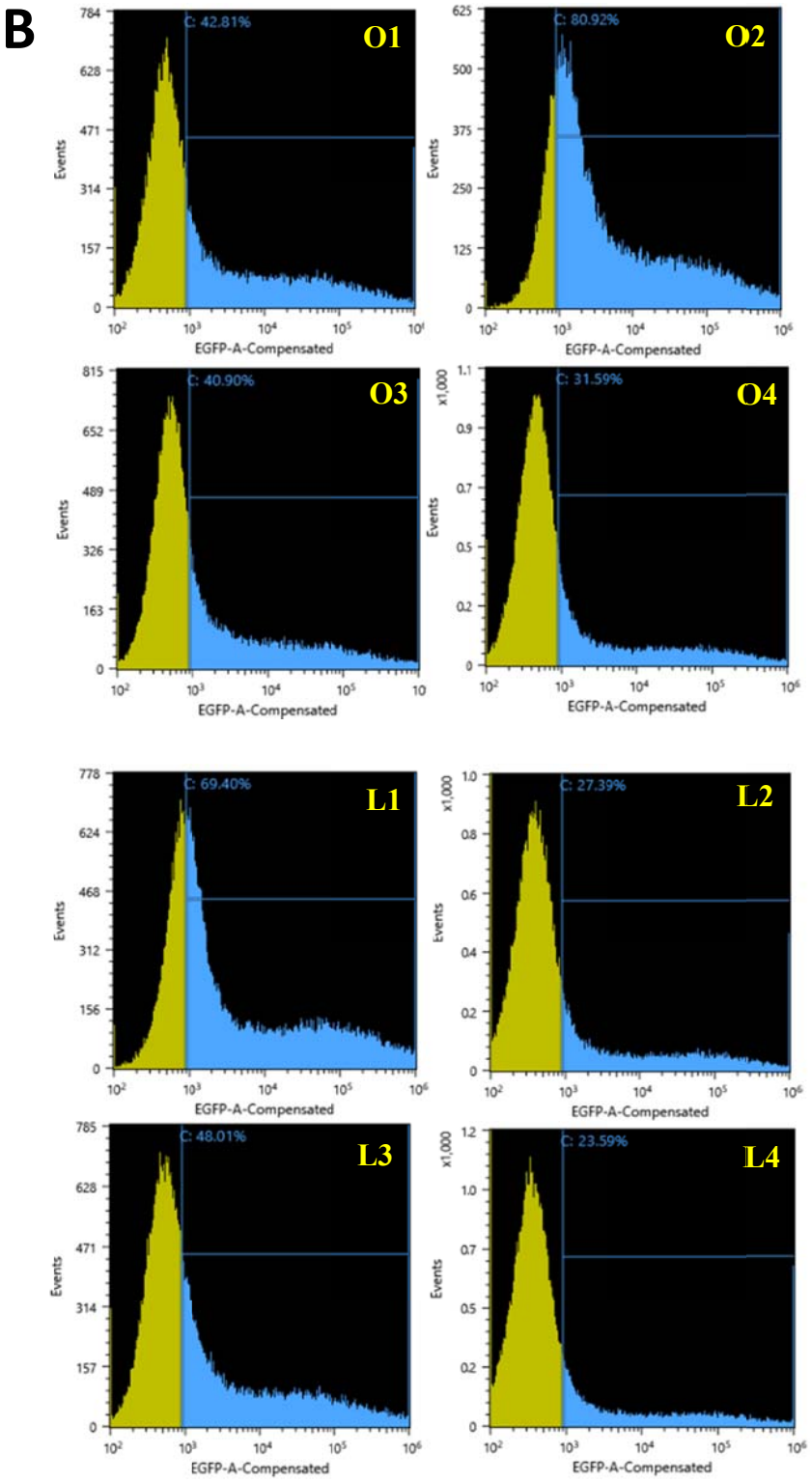


Fig. 3-3 SSA assay test gRNA sequences in *OVA* and *Lys* gene locus. (**A**) HEK293 cells fluorescence microscopy. O1–O4: gRNA/*OVA*1~gRNA/*OVA*4; L1~L4: gRNA/Lys1~gRNA/Lys4. (**B**) Evaluation of GFP-positive cells by FACS.

Table 3-4. SSA assay evaluation of GFP-positive cells by FACS

	1	2	3	4
gRNA/*OVA*	42.81%	80.92%	40.90%	31.59%
gRNA/Lys	69.40%	27.39%	48.01%	23.59%

3.3.2.3 T7E1 assay test gRNA KO efficiency

For further confirm cleavage efficiency, we amplified gRNA target regions from *OVA* and *Lys* gene to utilize T7E1 assay. T7E1 assay result showed in Fig. 3-4, no cleavage fragment in the T7EI assay could be obtained for all of the designed gRNAs. The *in vitro* cleavage efficiency was much lower than the predicted level, besides cell type, transfection efficiency, and gRNA off-target effects, chromosomal structure also affects the function of gRNA/Cas9 complexes on the specific DNA sequence. In eukaryotic cells, DNA strands wind around the nucleosome unit to form a high-order structure, and genes that should not be expressed are condensed to prevent their unregulated expression [164]. Approximately 146 bp of DNA intertwined in a left-handed supercoil turn embraces a histone octamer, consisting of two copies of the four core histone proteins (H2A, H2B, H3, and H4) [165], and a linker DNA with 60~80 bp is located adjacent to the nucleosome particles. When the target DNA sequence adopts a condensed chromatin structure, a wide range of DNA functional enzymes, such as restriction endonucleases and DNA repair enzymes, exhibit greatly reduced activity and efficiency [166, 167]. In the CRISPR/Cas9 system, the gRNA/Cas9 complex utilizes a three-dimensional diffusion-based mechanism in searching for PAM sites to slowly initiate DNA binding [168], and a 20-bp target DNA sequence is necessary for further recognition and binding [169-171]. In this process, Cas9 enzyme activity requires full access to the target DNA interface. However, it is suggested that, when the target DNA sequence is inside the nucleosome core, Cas9 endonuclease activity is inhibited by inaccessibility, while Cas9 activity is functional on the linker DNA sequence at the target DNA site. In this study, no cleavage fragment in the T7EI assay could be obtained for all of the designed gRNAs, suggested two gene 3'-UTR regions may inside of the nucleosome core.

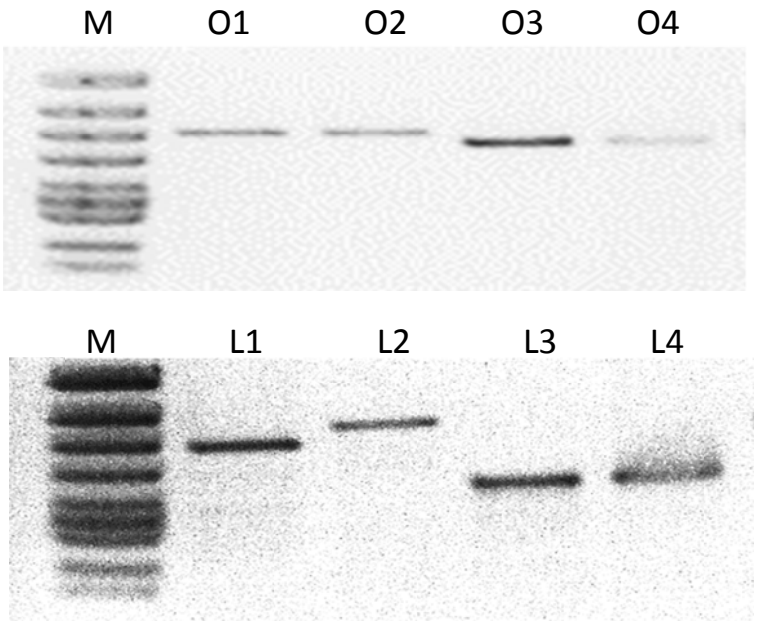


Fig. 3-4 gRNA/*OVA*1~4 and gRNA/Lys1~4 T7E1 assay result

3.3.2.4 Second round design gRNA location in *ovalbumin* gene and T7E1 assay test gRNA cleavage efficiency

To obtain high cleavage efficiency of gRNA sequence and search perhaps accessible region in ovalbumin gene, we redesigned gRNA sequences for whole introns and exons of the *OVA* gene, total 13 gRNA sequences, The location of redesigned gRNAs in ovalbumin gene was shown in Fig. 3-5, and the sequences used for gRNA were summarized in Table 3-1. After gRNA screening for the whole *OVA* region, the end region of the first intron resulted in an efficient gRNA, the region of which was assumed to be within the linker DNA or to contain a relaxed nucleosome structure, result showed in Fig. 3-6 and indel mutation rate was 31.1% by count.

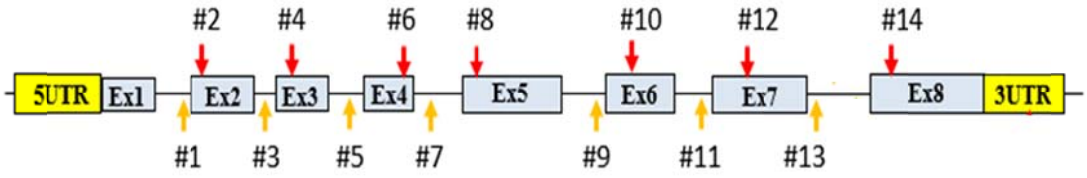


Fig. 3-5 The location of gRNA redesigned for transgene knock-in into the ovalbumin gene locus

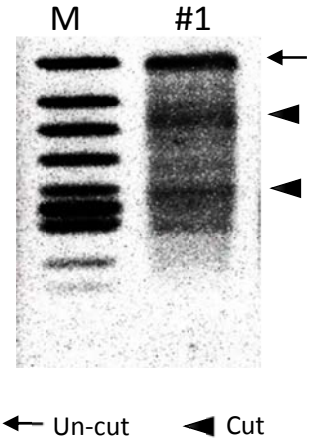


Fig. 3-6 gRNA/*OVA*-Intron1 in T7E1 assay result have expected cut band

3.4 Detect efficiency of HITI 1-cut donor vector to knock-in ovalbumin gene

3.4.1 Materials and Methods

3.4.1.1 Plasmid construction

A plasmid of pX330/*OVA*_i1 (gRNA/*OVA*-Intron1) for *OVA* intron 1 was prepared. A DNA fragment encoding P_{CMV}/*EGFP*/pA and P_{SV40}/Zeocin/pA expression units was amplified by PCR from pcDNA4/IRES-*EGFP* using the following primers: 5'-GGG TGC TGT ACA TAG TAC CAT GCA GGT TCG CGA TGT ACGGGC CAG A-3' and 5'-GGC AGT GAG CGC AAC GCA AT-3', which append a single digestion site for gRNA of *OVA* intron 1 (underlined) onto the 5' end of the product. PCR was initiated using KOD Plus Neo DNA polymerase (Toyobo, Osaka, Japan). PCR products were purified by NucleoSpin Gel and PCR Clean-up (MachereyeNagel, Düren, Germany), and then ligated into EcoRV-digested pBluescript II KS(＋) (Stratagene, La Jolla, CA, USA) to generate HITI donor vector pHITI/CMV-IRES-*EGFP*.

The schematic drawing of construction of pHITI/CMV-IRES-*EGFP* is showed in Fig. 3-7.

The sequences used for gRNA are summarized in Table 3-1.

3.4.1.2 Cell culture and transfection

The chicken fibroblast cell line UMNSAH/DF-1 (ATCC CRL-12203) were

cultured in highglucose D-MEM (SigmaeAldrich, St. Louis, MO, USA) supplemented with 10% (v/v) fetal bovine serum (BioWest, Nuaillé, France), 100 units/mL streptomycin sulfate (Fujifilm Wako Pure Chemical Industries, Osaka, Japan), and 100 mg/L penicillin G potassium (Fujifilm Wako) in tissue culture dishes (Thermo Fisher Scientific, Waltham, MA, USA). DF-1 cells were cultured at 39°C, in a 5% (v/v) CO_2 incubator.

For targeted knock-in, 1 day before transfection, DF-1 cells were seeded at a density of 1.2×10^6 cells per dish in 60-mm tissue culture dishes (Thermo Fisher Scientific) containing 5.0 mL of medium. The cells were co-transfected with 5 μg of donor vector and 5 μg of pX330-based Cas9/gRNA expression vectors using Lipofectamine 2000 (Invitrogen), in accordance with the manufacturer's instructions. After 2 days of culture, the cells were reseeded at a density of 1.0×10^5 cells/well in six-well tissue culture plates containing 2.0 mL of medium per well. Next day, drug screening was performed in 2.0 mL of medium containing 400 μg/mL Zeocin (Invitrogen) for 3 days. The medium was replaced with zeocin-containing fresh medium every other day. After 15 days of drug selection, colony number was counted to evaluate the efficiency of targeted knock-in. Cell clones were isolated by the colony picking method. Picked clones were subjected to flow cytometry analysis (SH-800; Sony, Tokyo, Japan) to observe *EGFP* expression.

3.4.1.3 Genomic PCR analysis

Genomic DNA was extracted from cells using a commercially available kit (MagExtractor Genome; Toyobo). Genomic DNA (50 ng) was subjected to PCR using DNA polymerase (KOD Plus Neo; Toyobo) at 94°C for 1 min; followed by 38 cycles of amplification at 95°Cfor 10 s, 56–60°C for 30 s, and 72°C for 20 s; and final extension at 72°C for 2 min. When the Tm value of primers was higher than 62°C, two-step PCR was performed (95°C for 10 s, 68°C for 30 s). The primers used are summarized in Table 3-5.

Table 3-5. Primers used for genomic PCR analysis in this study

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')[a]
α	TTCTCTTCTGTCTGAATGTCACCAC	
β		GCGGAACTCCATATATGGGC
γ	CCGCCGCCTTCTATGAAAGGTTG	
δ		ATACAATTAAGTGCTGTGGCTCC

3.4.2 Results and Discussion

To evaluate the HITI-mediated transgene knock-in system, we constructed a donor vector (pHITI/CMV-IRES-*EGFP*) containing *EGFP* and *Zeocin* expression cassettes as a reporter and drug selection marker, respectively, in which a complementary gRNA sequence was placed upstream of the transgene cassette, the schematic drawing of transgene (GOI) integration into a genomic locus by the HITI-mediated single-cut knock-in system was showed in Fig. 3-7A, and a schematic drawing of targeted knock-in of donor vector (pHITI/CMV-IRES-*EGFP*) into *OVA* intron 1 using pX330/*OVA*_i1 for the HITI-mediated CRISPR/Cas9 system was showed in Fig. 3-7B. DF-1 cells were co-transfected with pX330/*OVA*_i1 and the donor vector. Two days after the transfection, zeocin was added to the medium and the cells were cultured for 15 days. After cell clones showing zeocin resistance and *EGFP* expression had been isolated, genomic PCR analysis was performed for the cells to confirm 5' and 3' junctions of the knock-in regions, the electrophoresis result showed in Fig. 3-7C. For 20 clones exhibiting PCR-amplified fragments in both 5' and 3' junctions, the amplicons were applied for sequencing. Eleven out of 20 clones (55%) exhibited expected sequences in the 5' junction, but all 20 clones had mutations in the 3' junction, the sequencing result showed in Fig. 3-8. *EGFP* expression of the cell clones was further confirmed by flow cytometric analysis, indicating uniform expression levels for the clones, clone 1#, clone 3#, and clone 9# FACS result showed in Fig. 3-9. These results indicate that targeted knock-in of transgenes into the *OVA* locus was successfully achieved using the HITI-mediated transgene knock-in system.

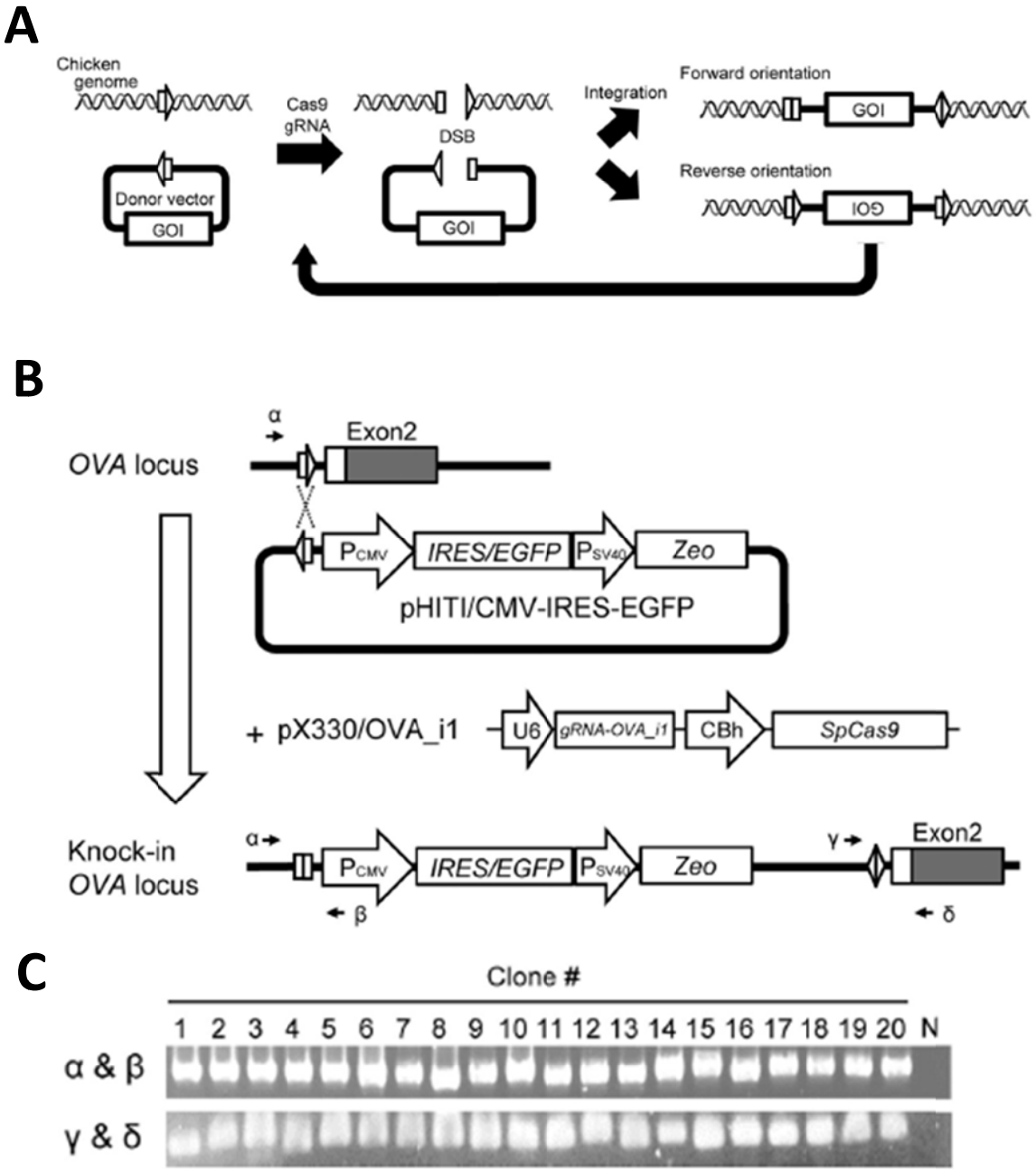


Fig. 3-7 Schematic of mechanism of HITI 1-cut knock-in system. (**A**) gRNA sequence in genome and donor vector were reverse complement, when Cas9 protein generated DSB, PAM (NGG) with subsequent 3 bp sequence (rectangle) in genome, linked the reverse complement PAM sequence of donor vector, same as the left 17 bp sequence (triangle) from genome, were connected to reverse complement sequence. When reverse orientation integration happened, which complete gRNA sequence retained in genome, Cas9/gRNA complex will recognize and cut gRNA sequence again until forward orientation integration or other kind form of absence complete gRNA sequence in genome, for example, random mutation in gRNA region. (**B**) Schematic of the donor vector (pHITI/CMV-IRES-*EGFP*) and Cas9/gRNA expression vector (pX330/*OVA*_i1) knock-in into genome using HITI system. Primers of α, β and γ, δ used for 5' and 3' junction genome PCR test for knock-in condition. *EGFP*, enhanced green fluorescent protein gene; P_{CMV}, human cytomegalovirus immediate early enhancer and promoter; IRES, internal ribosome entry site; P_{SV40}, simian vacuolating virus 40 promoter; $_{P}A$, polyA sequence; Amp^{R}, Ampicillin expression cassette; P_{U6}, avian U6 promoter for gRNA expression; hspCas9, a human codon-optimized SpCas9; Exon2, ovalbumin gene exon2. (**C**) Schematic of genomic PCR analysis. Electrophoresis result revealed both 20 clones were on-target knock-in.

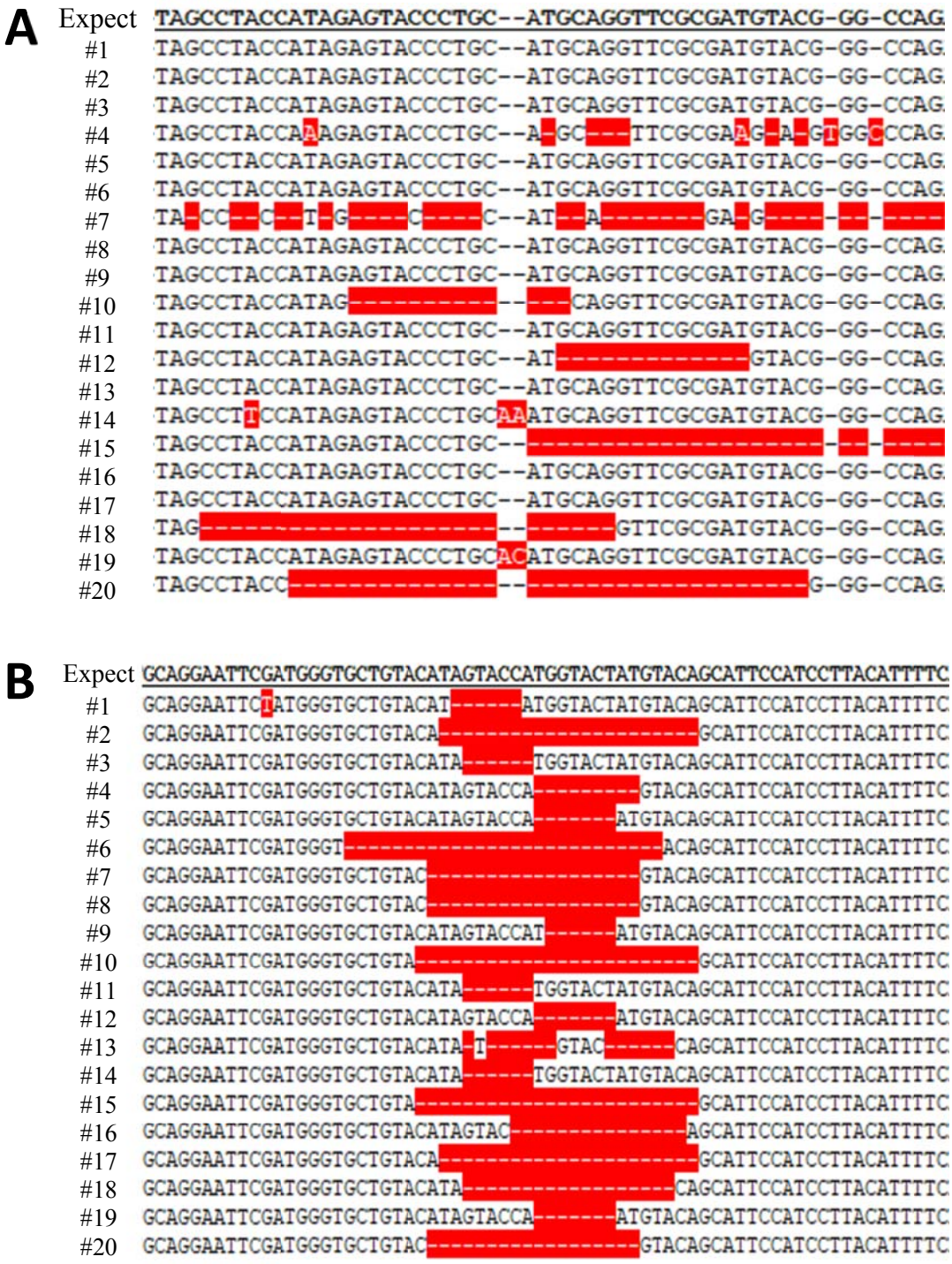


Figure 3-8 Sequencing result of 5' and 3' junctions 20 clones. (**A**) 5' junctions sequencing result, top sequence of refers to KI expected sequence. 10 clones were expectant integration. (**B**) 3' junctions sequencing result, top sequence of refers to KI expected sequence. All clones in 3' junctions has mutation.

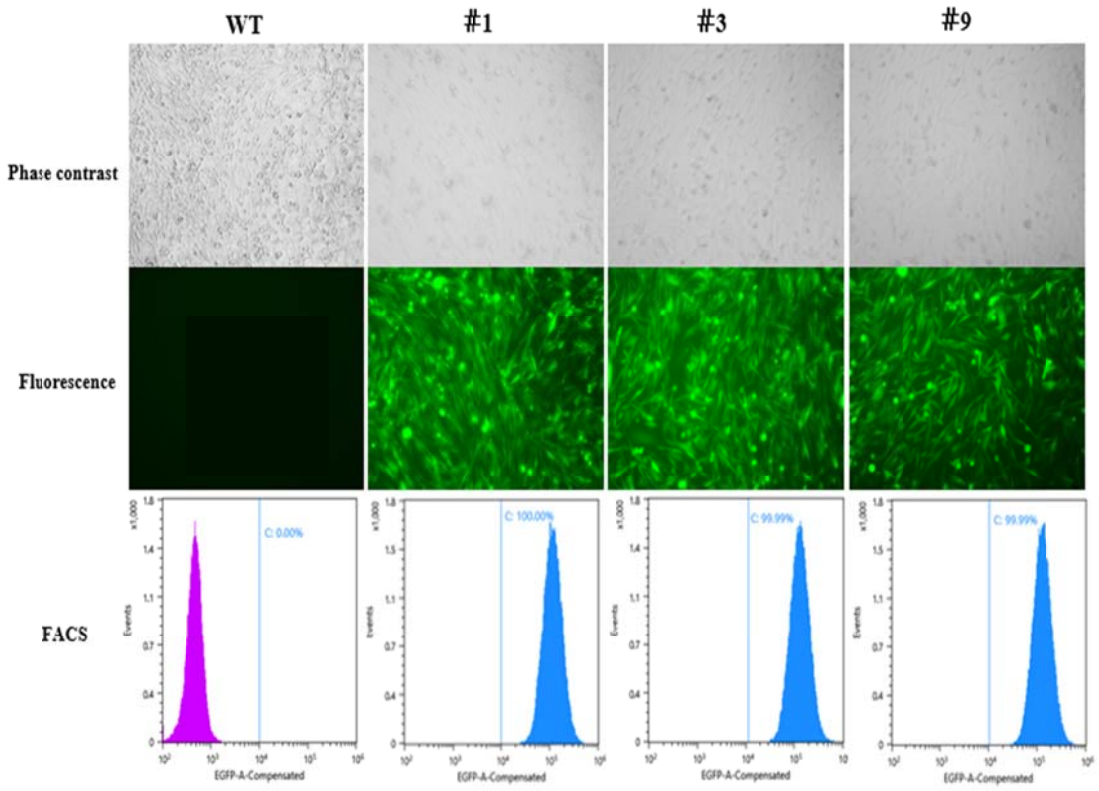


Figure. 3-9 Schematic of fluorescence microscope images and FACS result of *EGFP* expression in clone1#, clone3# and clone9#.

3.5 Conclusion

For transgene expression under control of the endogenous gene promoter, we designed four gRNA sequences of CRISPR/Cas9 around the *OVA* and *Lys* 3'-UTR region, so that the transgene was co-expressed with endogenous *OVA* or *Lys* via IRES or 2A peptides. Although we selected gRNA sequences exhibiting a high score in the CRISPRdirect software and high efficiency in the single strand annealing assay for evaluating gRNA digestion efficiency [172], the CRISPR vectors incorporated with the gRNA sequences could not work in CEF cells, as revealed by T7EI assay. To obtain high cleavage efficiency mediated by CRISPR/Cas9, we redesigned gRNA sequences for whole introns and exons of the *OVA* gene. A total of 13 gRNA sequences were introduced into the CRISPR vector, and CEF cells were transfected with the plasmids. After 48 h of transfection, cells were harvested. Then, genomic DNA was extracted from the cells to apply the T7EI assay to measure cleavage efficiency for each gRNA. The gRNA designed for intron 1 exhibited maximum

mutation efficiency of 31.1%. Therefore, this gRNA was used for further experiments.

The CRISPR/Cas9 system relies on cell repair mechanisms to maintain genomic integrity. Two major pathways, HDR and NHEJ, are involved in the repair of DSB. Although HDR is highly precise, it requires homology arms of 5' and 3' regions around the DSB, each of which should be 800 bp or more in length. Therefore, the construction of the HDR donor vector is time-consuming and laborintensive, and knock-in efficiency is lower than for NHEJ directly linking DSB [173]. Based on the NHEJ pathway, the HITI-mediated system uses one gRNA sequence to cleave target sites on the genome and donor vector [174]. After ligation of the two DSBs in the reverse direction, the sequence at the integration site is no longer recognized by gRNA because the original gRNA target sequence does not remain at the integration site. Thus, the direction of integration of a transgene is fixed. Another advantage of HITI is the low frequency of mutations. Although all clones in our study had mutations at the 3' junction, approximately 60% of the clones had error-free integration as expected at the 50 junction, with 100% targeted knock-in efficiency and no off-target effects. The reason for the high frequency of mutations at the 3' junction is unclear, and further research on the HITI-mediated integration mechanism is needed. A lower rate of mutation would occur when using minicircle donor.

Chapter 4

dCas9-VPR transactivation system initiating ovalbumin gene express in chicken cells

4.1 Introduction

Dead cas9 (dCas9), which has no nuclease activity, can fuse with different effector proteins. Under the guidance of gRNA, effector proteins can reach specific genomic sites to regulate the transcription of genes at specific sites, including CRISPRa and CRISPRi. At present, CRISPR/Cas9 system mediated transcription regulation has two main ways: one is to modify dCas9, the other is to modify gRNA. In the first way, the transcriptional activation is to connect a transcriptional activation effect protein on the dcas9 protein, which is directed to the target genomic region (usually the promoter position of the target gene or the coding region of the target gene close to ATG) by the gRNA, and the transcriptional activation effect protein will recruit more RNA polymerases, thus up regulating the transcriptional level of the gene; transcriptional inhibition It is to guide dcas9 to the target genome region (usually the coding region of the target gene) by gRNA, and inhibit the initiation and extension of transcription by physical blocking, so as to inhibit gene expression. The second mediated transcription regulation is to connect a small RNA structure on the gRNA, which binds to RNA binding protein, and RNA binding protein fuses transcription effector protein, so as to regulate the transcription level of the target gene.

4.2 Experimental Purpose

Egg white proteins such as *OVA* are specifically expressed in oviduct epithelial cells, which are difficult to obtain in a large number and difficult to culture for long periods of time. Thus, the study of egg white protein promoters has been a challenge. Considering the importance of ovalbumin gene in our research, selection of gRNA

sequence with dCas9 transactivation system initiating ovalbumin gene in non-oviduct cells is necessary. In the present study, we first designed a five gRNA sequences around the TATA box region of the *OVA* promoter. Then we test the efficiency of initiation *OVA* gene in DF-1 cells by dCas9 transactivation system.

4.3 Materials and Methods

4.3.1 Plasmid construction

Plasmid of pSLQ2814 (pSLQ/dCas9) expressing the fusion protein of SpdCas9-VPR, were obtained from Addgene (84247). Phosphorylated, and annealed in line with standard protocols. The gRNAs for the promoter region of the *OVA* gene were designed using CRISPRdirect software (http://crispr.dbcls.jp) [162]. After annealing, the gRNA templates were inserted into BbsI-digested pX330 (pX330/P1-pX330/P5 for *OVA* promoter regions). By digesting the plasmids (pX330/P1-pX330/P5) with PstI to remove the Cas9 expression unit, gRNA-expression vectors (pU6/P1-pU6/P5) were constructed.

Oligonucleotides for gRNA templates were chemically synthesized were showed in Table 4-1 (Invitrogen, Carlsbad, CA, USA)

Table 4-1. Sequences used for gRNA in dCas9 system

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')
pU6/dgRNA1[a,b]	CACCGtcaaaggtcaaacttctgaa	AAACttcagaagtttgacctttgaG
pU6/dgRNA2	CACCGttcttaaagatcccattatc	AAACgataatgggatctttaagaaG
pU6/dgRNA3	CACCGtcctgatggattagcagaac	AAACgttctgctaatccatcaggaG
pU6/dgRNA4	CACCGtgcacgttgtacatacaaga	AAACtcttgtatgtacaacgtgcaG
pU6/dgRNA5	CACCGaatgatttctatggcgtcaa	AAACttgacgccatagaaatcattG

[a] Target gRNA sequence is in lowercase.
[b] Modified px330 plasmids used for gRNA expression in dCas9 activation system.

4.3.2 Cell culture and transfection

The chicken fibroblast cell line UMNSAH/DF-1(ATCC CRL-12203) and

chicken embryonic fibroblasts (CEFs) derived from 12-day stage embryos (Line-M; Nisseiken, Yamanashi, Japan) were cultured in highglucose D-MEM (SigmaeAldrich, St. Louis, MO, USA) supplemented with 10% (v/v) fetal bovine serum (BioWest, Nuaillé, France), 100 units/mL streptomycin sulfate (Fujifilm Wako Pure Chemical Industries, Osaka, Japan), and 100 mg/L penicillin G potassium (Fujifilm Wako) in tissue culture dishes (Thermo Fisher Scientific, Waltham, MA, USA). DF-1 cells and CEFs were cultured at 39°C and 37°C, respectively, in a 5% (v/v) CO_2 incubator.

To induce endogenous *OVA* expression using the dCas9-VPR transactivation system, DF-1 cells and CEFs were seeded 1 day before transfection at a density of 1.2×10^6 cells/well in a six-well tissue culture plate, and co-transfected with each of pU6/P (pU6/P1–pU6/P5; 4 μg) and pSLQ/dCas9 (4 μg) using Lipofectamine 2000. When all five gRNAs were used for transfection, cells were transfected with pU6/P cocktails (0.8 μg of each pU6/P1–pU6/P5; total of 4 μg) and pSLQ/dCas9 (4 μg). When two gRNAs (pU6/P1 and pU6/P5) were used for transfection, cells were transfected with pU6/P1 and pU6/P5 (2 μg each) and pSLQ/dCas9 (4 μg). At 48 h post-transfection, the cells were analyzed for further experiments.

4.3.3 Isolation of total mRNA and qRT-PCR analysis

Total mRNAs from transfected DF-1 and CEF were isolated using a kit (RNAiso Plus; Takara Bio, Kusatsu, Japan), in accordance with the manufacturer's protocol. The RNA samples were reverse-transcribed into cDNA using a kit (ReverTra Ace First Strand cDNA synthesis kit; Toyobo) with oligo-dT primers. RT-PCR was initiated with KOD Plus Neo DNA polymerase (Toyobo) at 94°C for 1 min; 38 cycles of amplification at 95°C for 10 s, 58°C for 30 s, and 72°C for 20 s; and final extension at 72°C for 2 min. mRNA quantification was performed using Thunderbird SYBR qPCR Mix (Toyobo) and AriaMx real-time PCR system (Agilent Technologies, Santa Clara, CA, USA). The fold change in the *OVA*-specific transcripts relative to the GAPDH housekeeping control gene was determined by the △△Ct method. The mRNA expression levels are expressed as mean and standard deviation. The primer

pairs used for *OVA* gene amplified and qRT-PCR are summarized in Table 4-2.

Table 4-2. The primer pairs used for *OVA* gene was amplified and qRT-PCR

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')
OVA-PCR	TGTGGCACATCTGTAAACGTTCACT	CATTTGTCTGACTTTCTACCCAGG
OVA-qPCR	CTCAACCAAATCACCAAACCAAA	TGCAGCTGTTTGAAAGTTGATA
GAPDH-qPCR	CACAGACGGTGGATGGC	GAGACATTGGGGGTTGGCA

4.3.4 Western blot analysis

Reducing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by western blotting was performed to detect *OVA* protein induced using the dCas9-VPR activation system in transfected DF-1 cells. Samples were prepared from DF-1 cells transfected with pU6/P cocktails (pU6/P1–pU6/P5) and pSLQ/dCas9 with 1 mL of cell lysis reagent (RIPA lysis and extraction buffer; Invitrogen), in accordance with the manufacturer's protocol. After cell lysis solution had been mixed with SDS-PAGE sample buffer with bmercaptoethanol and boiled at 100°C for 5 min, the supernatant was subjected to SDS-PAGE on a pre-cast 4%e12% NuPAGE minigel (Invitrogen) with 2-(Nmorpholino) ethanesulfonic acid (MES) buffer. Samples derived from untransfected cells and 100 ng of ovalbumin (Fujifilm Wako) were also used as controls. After electrophoresis and transfer onto polyvinylidene fluoride membranes using the iBlot Gel transfer system (Invitrogen), the membranes were immersed in Trisbuffered saline containing 0.02% (v/v) Tween20 and 5% (w/v) nonfat milk at 4°C overnight for blocking. *OVA* proteins on the membranes were detected using rabbit anti-*OVA* antibody (Novus Biologicals, Centennial, CO, USA) at 1:1000 dilution and a peroxidase (POD)-conjugated goat anti-rabbit IgG secondary antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) at 1:1500 dilution. Can Get Signal solution (Toyobo) was used in the antibody dilution to enhance signal detection. The specific antibodyeantigen complexes were detected using an enhanced chemiluminescence kit (GE Healthcare, Waukesha, WI, USA).

4.4 Results and Discussion

4.4.1 Each gRNA sequence activation efficiency in *OVA* gene promoter region with dCas9 system

To activate endogenous *OVA* gene expression, the dCas9-VPR transactivation system was employed. First, five gRNA sequences around the TATA box of the *OVA* promoter were designed, gRNA location around TATA box of ovalbumin gene promoter were showed in Fig. 4-1. DF-1 and CEF cells were transfected with the dCas9-VPR expression vector together with gRNA expression vectors encoding the respective gRNA sequences (pU6/P1~pU6/P5). RNA was extracted from the cells on day 3 after transfection, and *OVA* gene expression was evaluated by RT-PCR using specific primers. Using gRNA-P1 and gRNA-P5, clear amplified bands could be obtained. The sequence analysis of amplicons revealed that endogenous *OVA* gene expression was induced for both DF-1 and CEF cells, electrophoresis result of RT-PCR showed in Fig. 4-2, the sequence analysis of gRNA-P1 and gRNA-P5 amplicons showed in Fig. 4-3. The induction level of gRNA-P1 was 5.5- and 11-fold higher than that of gRNA-P5 for DF-1 and CEF cells, respectively, qRT-PCR result showed in Fig. 4-4. When all of the five gRNAs were used, *OVA* gene expression was further enhanced.

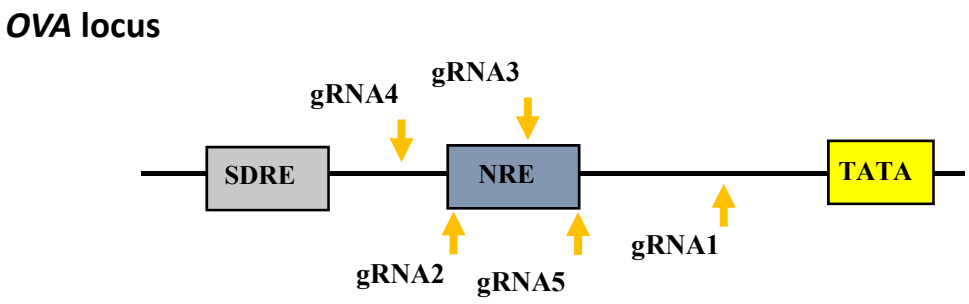


Fig. 4-1 Schematic of dCas9 activation system structure and gRNA locations in *OVA* gene prompter region. gRNA1 and gRNA5 near TATA box, gRNA2 and gRNA3 in NRE region, gRNA4 inside of NRE and SDRE

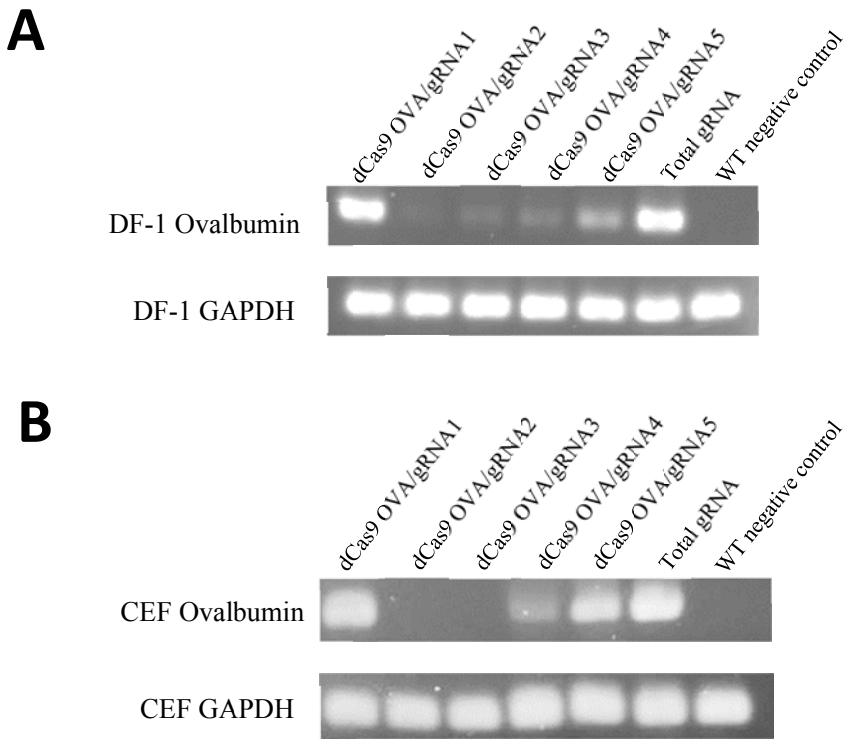


Fig. 4-2 ovalbumin mRNA expression in DF-1 (**A**) and CEF cells (**B**) after dCas9 system transfected with single or total gRNAs. gRNA3 and gRNA4 had slight bands in RT-PCR electrophoresis.

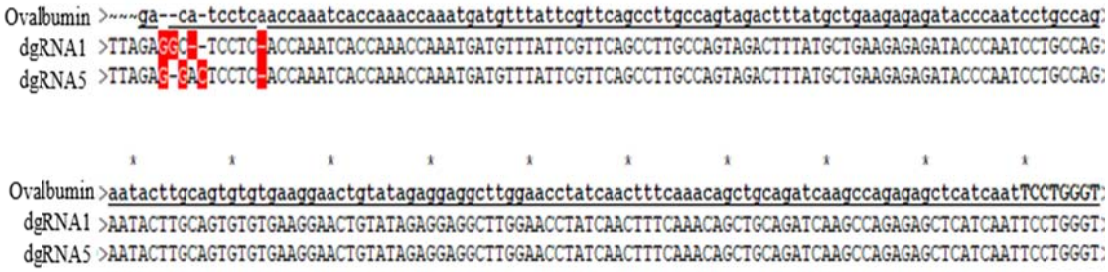


Fig. 4-3 Sequence alignment: Top sequence refers to ovalbumin cDNA sequence after amplified which are 247bp

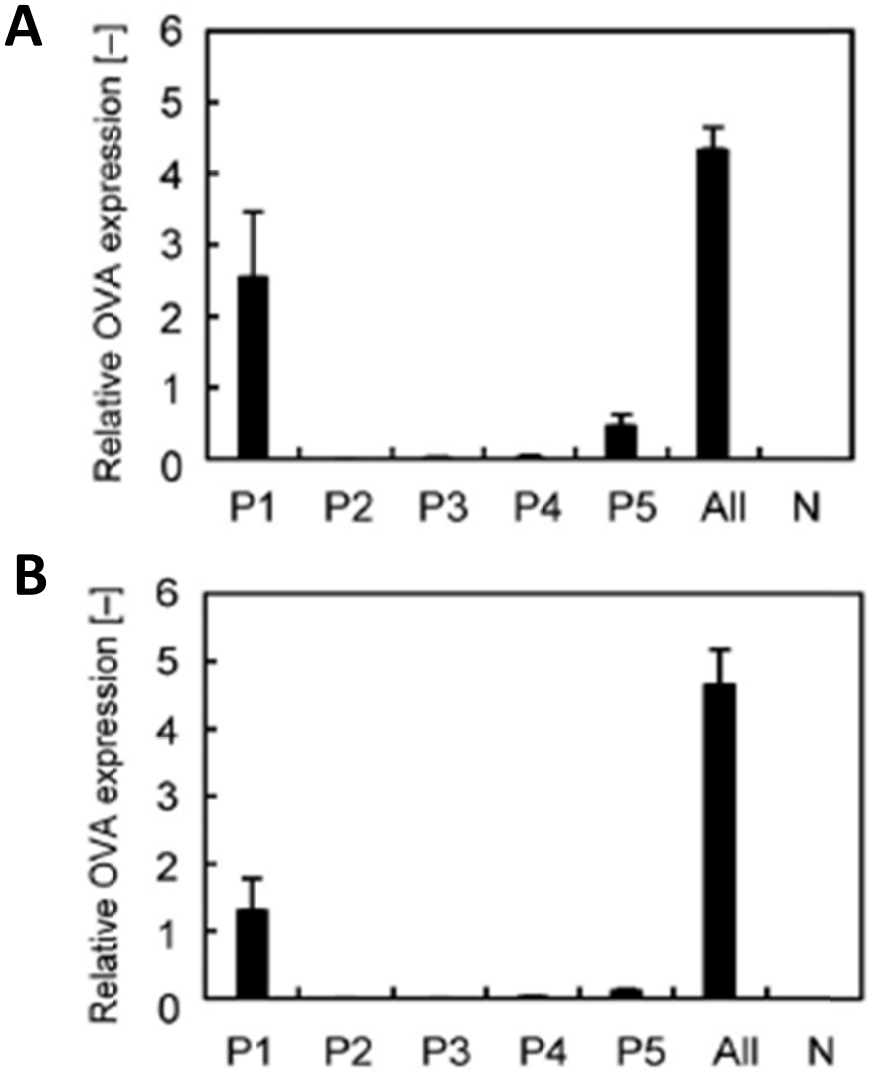


Fig. 4-3 In qPCR analysis, gRNA 1 was more efficiency than gRNA5, with total 5 gRNA co-transfection, both in DF-1 (**A**) and CEF cells (**B**) got highest mRNA expression level. P1~P5, single gRNA was used; All, all five gRNAs were used; N, no gRNA.

4.4.2 Detect different gRNA combination activate gene expression efficiency

To confirm that the combination of gRNA-P1/gRNA-P5 is sufficient to induce the maximum level of *OVA* expression, endogenous *OVA* gene induction by the dCas9-VPR transactivation system was evaluated with comparison between the combination of gRNA-P1/gRNA-P5 and all five gRNAs using DF-1 and CEF cells. Total mRNAs extracted from the cells at 3 days after transfection were applied for RT-PCR and qRT-PCR. The results which showed in Fig. 4-4 and Fig. 4-5 revealed that the induction level of the endogenous *OVA* gene was higher for all five gRNAs

compared with the combination of gRNA-P1/gRNA-P5.

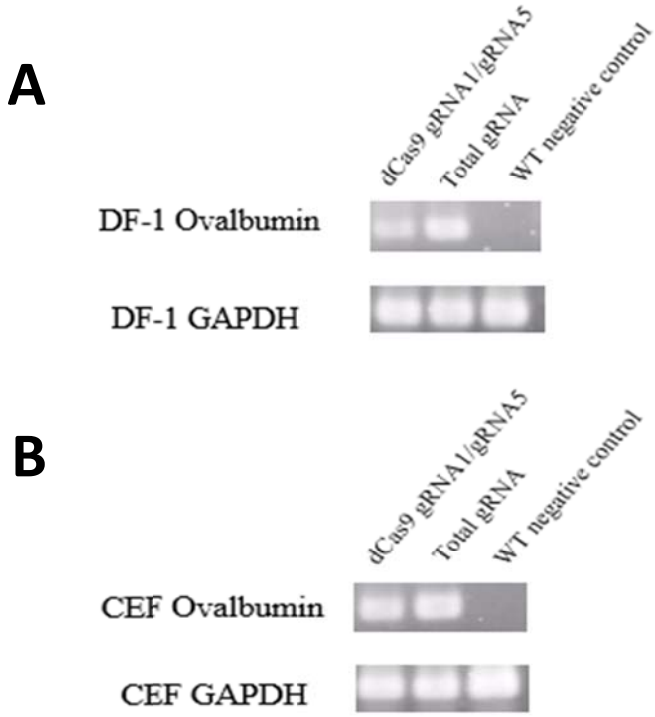


Fig. 4-4 ovalbumin mRNA expression in DF-1 (**A**) and CEF cells (**B**) after dCas9 system transfected with gRNA1 with gRNA5 or total gRNAs.

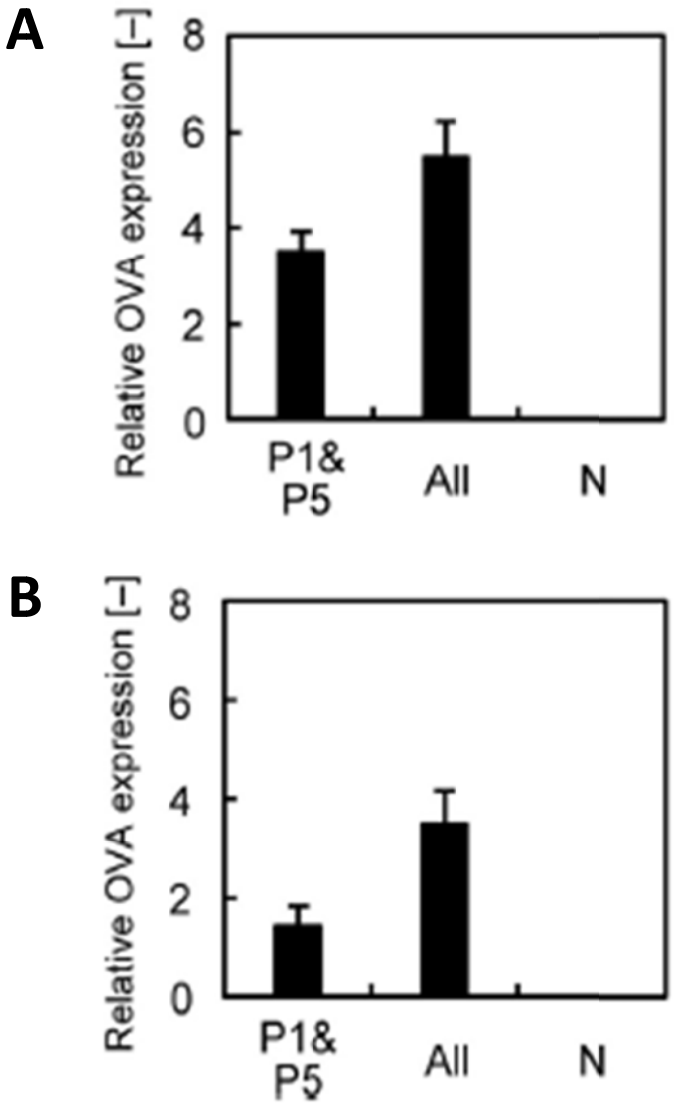


Fig. 4-5 In qPCR analysis, contrast with co-transfection of gRNA1 and gRNA5, total 5 gRNAs with dCas9 system still have higher efficiency in gene activation in DF-1 (**A**) and CEF cells (**B**). All means all five gRNAs were used; P1&P5 means gRNA-P1 and gRNA-P5 were used; N means no gRNA.

4.4.3 dCas9 activation system initiating ovalbumin protein expression in DF-1 cell

To further illustrate endogenous *OVA* gene induction using the dCas9-VPR transactivation system, *OVA* protein expression was confirmed by Western blot analysis, result showed in Fig. 4-6. A band corresponding to ovalbumin appeared with the expected size. These results indicated that the dCas9-VPR transactivation system using five gRNAs near the TATA box of the *OVA* promoter could effectively initiate endogenous *OVA* gene expression in non-oviduct cells.

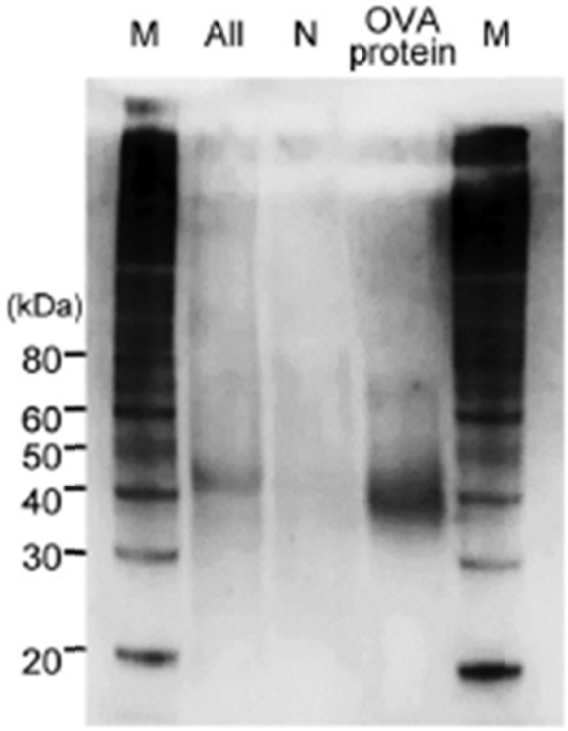


Fig. 4-6 Western blot test ovalbumin protein expressing in DF-1 cell after dCas9 system transfection with total 5 gRNAs. 20 ul of cell lysis which 5 gRNA and dCas9-expression vector co-transfected was detected band around 45kDa, DF-1 cell without operation set as negative control, positive control of 100 ng egg white protein around 40kDa had clear band. All, all five gRNAs were used; N, no gRNA, M, protein marker.

4.5 Conclusion

Egg white proteins such as ovalbumin are specifically expressed in oviduct epithelial cells, which are difficult to obtain in a large number and difficult to culture for long periods of time [175, 176]. Thus, the study of egg white protein promoters has been a challenge. In recent years, gene-specific gRNA-guided catalytically inactive Cas9 (dCas9) fusion proteins combining effector domains to target promoter DNA sequences have been widely used to regulate the transcription of endogenous genes [177]. We designed five gRNA sequences around the TATA box region of the *OVA* promoter. The gRNA-guided dCas9-VPR was successfully recruited to the *OVA* promoter without cleavage of the genome and initiated gene transcription. Our study indicated that steric hindrance of the dCas9-VPR protein and the spatial distance between PAM and TATA box sequences might be very important in the dCas9-VPR

transactivation system. In this study, gRNA-P1 and gRNA-P5 exhibited high efficiency in initiating gene expression, and the distances between PAM and TATA box sequences were 27 and 43 bp, respectively, while those of other gRNAs were more than 100 bp. Interestingly, transfection of all five of the gRNAs exhibited better transactivation activity than the transfection of gRNA-P1 and gRNA-P5, although gRNAs other than gRNA-P1 and gRNA-P5 did not show transcriptional activation activity using dCas9-VPR. We speculated that the gRNA/dCas9 complex could alter chromosomal structures, leading to target sites of the closely associated nucleosome becoming accessible.

Chapter 5

Induction of transgene expression under control of endogenous *OVA* promoter using dCas9-VPR transactivation system

5.1 Introduction

In the 1970s, researchers began to study the regulatory sequence of ovalbumin gene expression. In 1979, Royal et al. cloned 46 KB chicken DNA fragment, including X gene, Y gene and ovalbumin gene [178]. In 1982, Sanders et al. [55] carried out detailed research and Analysis on the region of ovalbumin gene-1340 ~ + 392, and determined the transcription starting site of ovalbumin gene according to its leading exon region and Tata frame of 5 'flanking region. It was found that there were more Tata frame like sequences at –870 ~ –700, which played an important role in the tissue-specific expression of chicken. In 1982, they further analyzed several regulatory elements through exclusion. In 1995, Haecker et al. [58] studied SRE and NRE elements in detail on the basis of predecessors, and found that there were hormone dependent regulatory elements (SDRE) in –872 ~ –780 region, which can respond to estrogen. -There is a group of trans-regulatory factors in 308-88 region, which contains three suppressors. In 1995, Park et al. [179] used the primary chicken oviduct epithelial cells to verify that the expression of –87 ~ +9 region of ovalbumin regulatory region (coup) was not enhanced, but it could change the estrogen dependence of SDRE region; Lillico [180] used 675 bp ere element and about 2.9 kb ovalbumin promoter to drive exogenous protein to obtain tissue-specific protein in oviduct Sexual expression. Byun et al. [181] constructed a lentivirus vector with about 2 kb ovalbumin promoter to drive the expression of e*EGFP* in tubal gland cells. Kaleri et al. [182] realized the specific expression of exogenous genes in fallopian tube by using 5 'upstream 2.8 kb fragment (including 5' upstream 1.2 kb fragment, the

first exon and the first intron); in 2018, Kwon et al. [183] also successfully initiated the specific expression of human epidermal growth factor in chicken fallopian tube by using ere estrogen response element combined with 1179bp fragment covering SDRE and NRE regions Heterosexual expression.

Ovalbumin promoter has long been considered as one of the strongest promoters in the body. With the continuous in-depth research of researchers, the regulatory elements that can play the startup function include ere, SDRE and NRE elements. These research results provide a reference for the selection of specific promoters used in this study.

5.2 Experimental Purpose

In recent years, gene-specific gRNA-guided catalytically inactive Cas9 (dCas9) fusion proteins combining effector domains to target promoter DNA sequences have been widely used to regulate the transcription of endogenous genes. Considering the importance of *OVA* gene promoter in our research, identify knock-in transgene could be regulated by endogenous *OVA* promoter is necessary. In the present study, we first prepare DF-1 cell line the knock-in *EGFP* cassette without *CMV* promoter, after check KI situation, then co-transfected five gRNA sequences around the *TATA* box with dCas9-VPR transactivation system to initiate transgene *EGFP* express.

5.3 Prepare transgene *EGFP* knock-in DF-1 cell line under endogenous *OVA* promoter

5.3.1 Materials and Methods

5.3.1.1 Plasmid construction

Plasmid of pX330/*OVA*_i1 for *OVA* intron 1 was constructed as section 3.2.2 materials and methods, A DNA fragment encoding PSV40/Zeocin/pA expression units was amplified by PCR from pcDNA4/IRES-*EGFP*, primers sequence showed in Table 5-1. PCR was initiated using KOD Plus Neo DNA polymerase (Toyobo, Osaka, Japan). PCR products were purified by NucleoSpin Gel and PCR Clean-up

(MachereyeNagel, Düren, Germany), and then ligated into EcoRV-digested pBluescript II KS(＋) (Stratagene, La Jolla, CA, USA) to generate HITI donor vector pHITI/CMV-IRES-*EGFP*. After digestion with NruI and EcoRV in pHITI/CMV-IRES-*EGFP* to delete the CMV promoter, pHITI/IRES-*EGFP* was generated.

Table 5-1. Primer sequences used for vector construction.

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')
Intron1-1cut[a]	GGGTGCTGTACATAGTACCATGCAGGTTCGCGATGTACGGGCCAGA	GGCAGTGAGCGCAACGCAAT

[a] gRNA/Intron1 sequence is underline.

5.3.1.2 Cell culture and transfection

The chicken fibroblast cell line UMNSAH/DF-1(ATCC CRL-12203) were cultured in highglucose D-MEM (SigmaeAldrich, St. Louis, MO, USA) supplemented with 10% (v/v) fetal bovine serum (BioWest, Nuaillé, France), 100 units/mL streptomycin sulfate (Fujifilm Wako Pure Chemical Industries, Osaka, Japan), and 100 mg/L penicillin G potassium (Fujifilm Wako) in tissue culture dishes (Thermo Fisher Scientific, Waltham, MA, USA). DF-1 cells and CEFs were cultured at 39°C, in a 5% (v/v) CO_2 incubator.

For targeted knock-in, 1 day before transfection, DF-1 cells were seeded at a density of 1.2×10^6 cells per dish in 60-mm tissue culture dishes (Thermo Fisher Scientific) containing 5.0 mL of medium. The cells were co-transfected with 5 μg of donor vector and 5 μg of pX330-based Cas9/gRNA expression vectors using Lipofectamine 2000 (Invitrogen), in accordance with the manufacturer's instructions. After 2 days of culture, the cells were reseeded at a density of 1.0×10^5 cells/well in six-well tissue culture plates containing 2.0 mL of medium per well. Next day, drug screening was performed in 2.0 mL of medium containing 400 mg/mL zeocin (Invitrogen) for 3 days. The medium was replaced with zeocin-containing fresh medium every other day. After 15 days of drug selection, colony number was counted to evaluate the efficiency of targeted knock-in. Cell clones were isolated by the

colony picking method.

5.3.1.3 Genomic PCR analysis

Genomic DNA was extracted from cells using a commercially available kit (MagExtractor Genome; Toyobo). Genomic DNA (50 ng) was subjected to PCR using DNA polymerase (KOD Plus Neo; Toyobo) at 94°C for 1 min; followed by 38 cycles of amplification at 95°C for 10 s, 56–60°C for 30 s, and 72°C for 20 s; and final extension at 72°C for 2 min. When the Tm value of primers was higher than 62°C, two-step PCR was performed (95°C for 10 s, 68°C for 30 s). The primers used are summarized in Table 5-2.

Table 5-2. Primers used for genomic PCR analysis in this study

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')[a]
α	TTCTCTTCTGTCTGAATGTCACCAC	
γ	CCGCCGCCTTCTATGAAAGGTTG	
δ		ATACAATTAAGTGCTGTGGCTCC
ε		CAAGCGGCTTCGGCCAGTAA

5.3.1.3 Off-target analysis

Off-target effects were predicted using the CRISPRdirect software (http://crispr.dbcls.jp) to identify potential off-target sites for *OVA* locus-specific gRNA. Genomic regions around each candidate site were amplified by PCR using the primer pairs showed in Table 5-3, and PCR amplicons were analyzed by direct sequencing.

Table 5-3. Primers used for off-target PCR analysis in this study

Name	Fw Sequences (5'→3')	Rv Sequences (5'→3')[a]
Intron1 off-target1	TGGCAATGCAACATGGCATT	GCTGAATATGCCACTACCTCCC
Intron1 off-target2	GGGAGACTTCAGTACAAAGTG	GGTGATGTAGTTTGAGCTCCC

5.3.2 Results and Discussion

5.3.2.1 Transgenes integration using 1 cut HITI donor vector

To determine whether the endogenous *OVA* promoter can initiate expression of

an exogenous gene integrated using the HITI-mediated transgene knock-in system, we constructed a donor vector encoding an *EGFP* expression cassette without the *CMV* promoter, for targeted knock-in into the endogenous *OVA* locus, a schematic drawing of targeted knock-in of donor vector (pHITI/IRES-*EGFP*) into *OVA* intron 1 using pX330/*OVA*_i1 for the HITI-mediated CRISPR/Cas9 system was showed in Fig. 5-1A. DF-1 cells were co-transfected with the donor vector and Cas9/gRNA vector (pX330/*OVA*_i1), and knock-in cells were screened using drug selection. Subsequently, targeted integration of the transgene was evaluated by genomic PCR using specific primer pairs. A total of 18 established clones were subjected to PCR analysis at the 5' and 3' junction regions, and DNA fragments with the expected sizes appeared for all clones, the electrophoresis result showed in Fig. 5-1B. Sequence analysis of the amplicons revealed that the samples from 12 out of 18 clones showed perfect integration in the 5' junction (67%), and that all samples showed mutations in the 3' junction, result showed in Fig. 5-2.

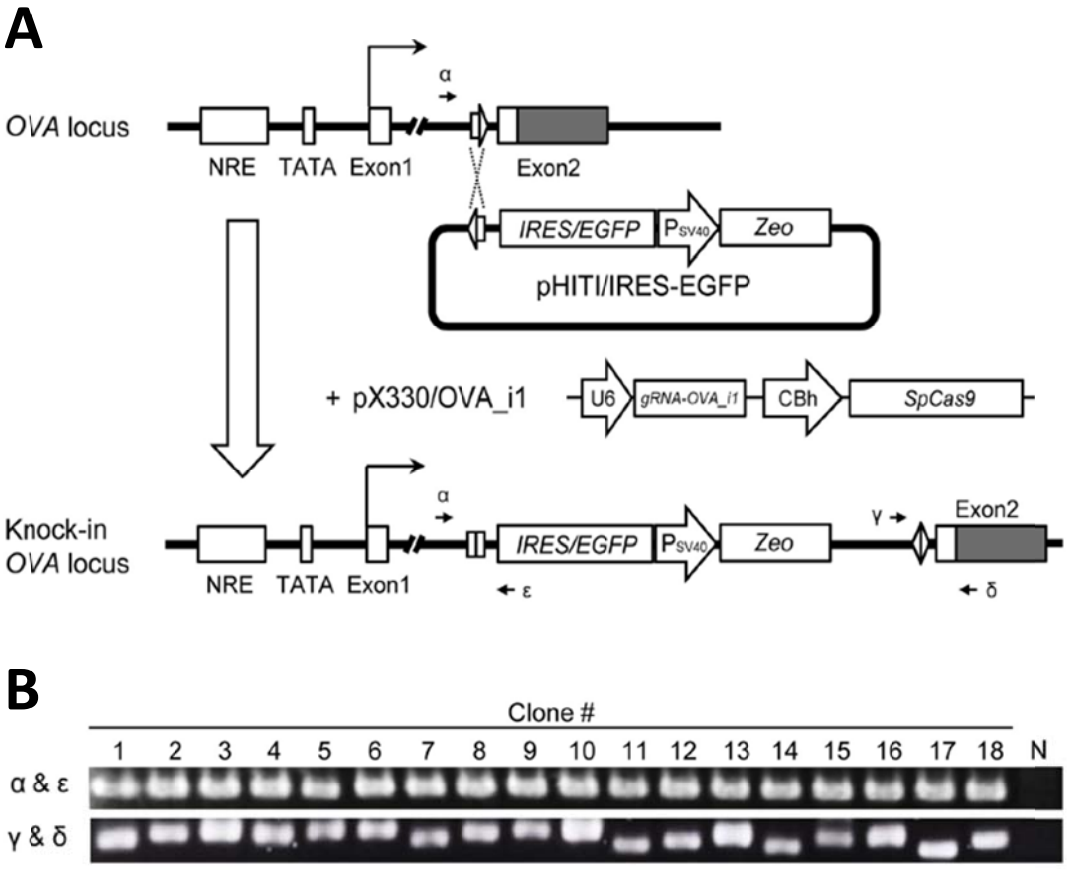


Figure. 5-1 dCas9 system activate *EGFP* gene expression under initiating *OVA* gene promoter in *EGFP/CMV*-delete KI clone cells which donor vector delete *CMV* promoter. (**A**) Schematic of under dCas9 activation, *EGFP* was initiated by *OVA* promoter. VP64, P16, RTA, transcriptional activators. Primers of α, ε and γ, δ used for 5' and 3' junction genome PCR test for knock-in condition. *EGFP*, enhanced green fluorescent protein gene; IRES, internal ribosome entry site; P_{SV40}, simian vacuolating virus 40 promoter; $_{P}A$, polyA sequence; Amp^{R}, Ampicillin expression cassette; P_{U6}, avian U6 promoter for gRNA expression; hspCas9, a human codon-optimized SpCas9; Exon2, ovalbumin gene exon2. (**C**) Schematic of genomic PCR analysis. Electrophoresis result revealed both 18 clones were on-target knock-in.

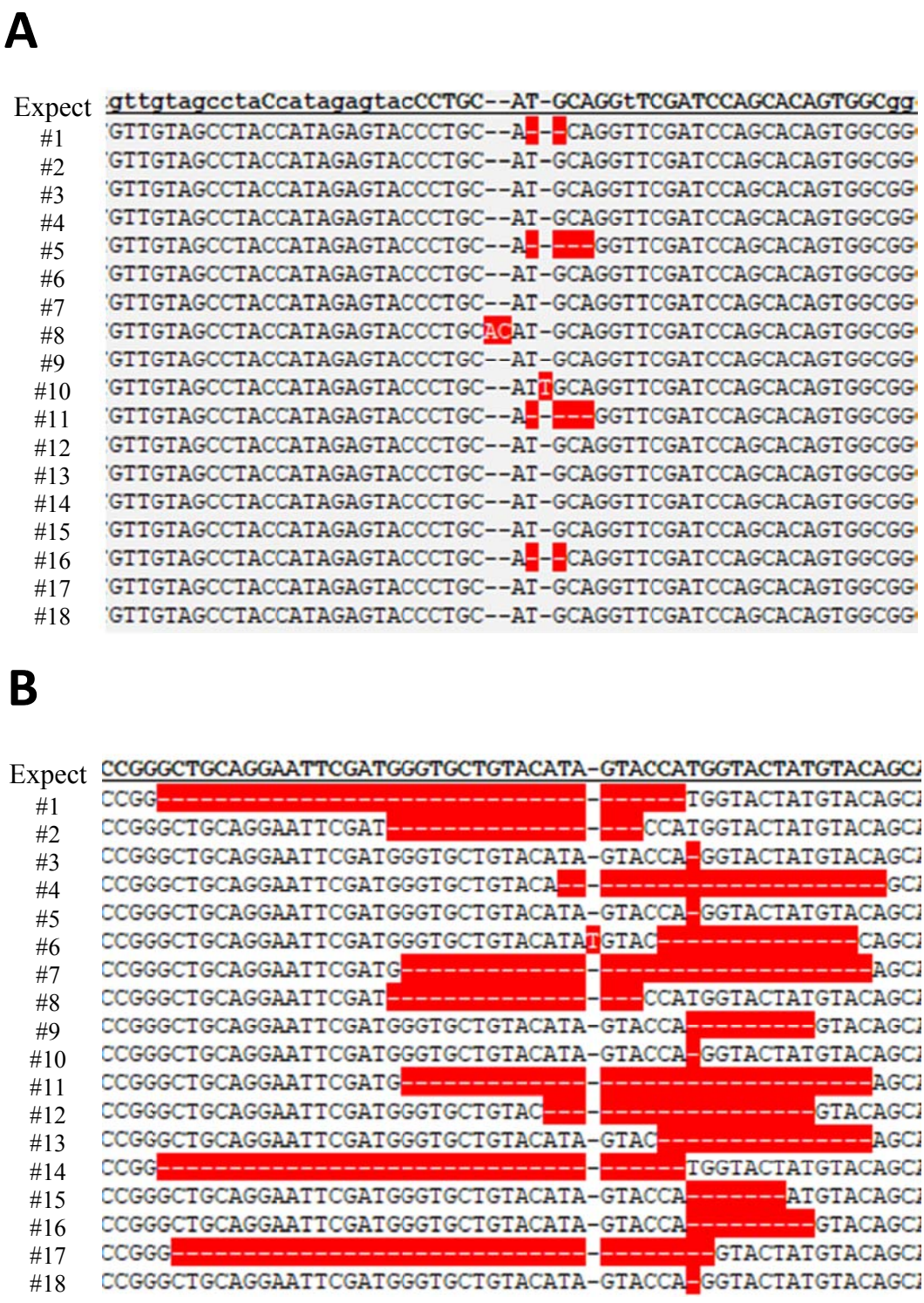


Figure. 5-2 Sequencing result of 5' and 3' junctions of 18 clones. (**A**) 5' junctions sequencing result, top sequence of refers to KI expected sequence, 12 clones were expectant integration. (**B**) 3' junctions sequencing result. All clones in **3**' junctions has mutation.

5.3.2.2 Off-target analysis in two kinds of KI DF-1 cell lines

To determine gRNA/*OVA*-Intron1 off-target situation, we predicted 2 off-target

sites by using the CRISPRdirect software. We choose each 10 clones from DF-1 cell line which using donor vector pHITI/CMV-IRES-*EGFP* and pHITI/IRES-*EGFP* for genomic PCR analysis. Genomic regions around each candidate site were amplified by PCR, and PCR amplicons were analyzed by direct sequencing, result showed in Table 5-4. There was no off-target in total 20 clones.

Table 5-4. Off-target analysis in DF-1 knock-in cell clones

		Sequence[a]	*EGFP* KI clones	*EGFP/CMV*-delete KI clones
***OVA* intron1**	NW_003763673.1: 87428762~87428785	**CCT**GCATGGTAC TATGTACAGCA		
off-target 1	NW_001488830.2: 9916838~9916861	**CCT**GCATGGTAC TATAAAATAAT	0/10	0/10
off-target 2	NW_003763487.1: 75218772~75218795	TTACAGTTATAGT ACCATGC**AGG**	0/10	0/10

[a] PAM sequence is written overstriking, mismatches are indicated by underling

5.4 Induction of *EGFP* expression under control of *OVA* promoter using dCas9-VPR transactivation system

5.4.1 Materials and Methods

5.4.1.1 Plasmid construction

Plasmid of pSLQ2814 (pSLQ/dCas9) expressing the fusion protein of SpdCas9-VPR, were obtained from Addgene (84247). Phosphorylated, and annealed in line with standard protocols. The gRNAs for the promoter region of the *OVA* gene were designed using CRISPRdirect software (http://crispr.dbcls.jp) [162]. After annealing, the gRNA templates were inserted into BbsI-digested pX330 (pX330/P1-pX330/P5 for *OVA* promoter regions). By digesting the plasmids (pX330/P1-pX330/P5) with PstI to remove the Cas9 expression unit, gRNA-expression vectors (pU6/P1–pU6/P5) were constructed.

5.4.1.2 Cell culture and transfection

To induce endogenous *OVA* expression using the dCas9-VPR transactivation

system, *EGFP* knocked-in DF-1 cells were seeded 1 day before transfection at a density of 1.2×10^6 cells/well in a six-well tissue culture plate, and co-transfected with all five gRNAs pU6/P cocktails (0.8 μg of each pU6/P1–pU6/P5; total of 4 μg) and pSLQ/dCas9 (4 μg) using Lipofectamine 2000, in accordance with the manufacturer's instructions. At 48 h post-transfection, the cells were subjected to flow cytometry analysis (SH-800; Sony, Tokyo, Japan) to observe *EGFP* expression.

5.4.2 Results and Discussion

To induce *EGFP* expression by the endogenous *OVA* promoter, a dCas9-VPR transactivation system using the five gRNAs effective for *OVA* expression was introduced into the knock-in cell clones (no. 3, 9, and 10). Three days after transfection of the dCas9-VPR and gRNA vectors, cells were observed using a fluorescence microscope and analyzed by flow cytometry, result showed in Fig. 5-3. Green fluorescent cells were detected for the clones. The frequencies of green fluorescent cells for clones no. 3, 9, and 10 were 7.4%, 5.0%, and 6.7%, respectively. The proportion of green fluorescent cells was low, reflecting the transfection efficiency of the vectors, because all vectors should be introduced into cells. Nevertheless, these results indicated that endogenous *OVA* promoter was able to initiate exogenous gene expression of the knock-in cells using the dCas9-VPR transactivation system.

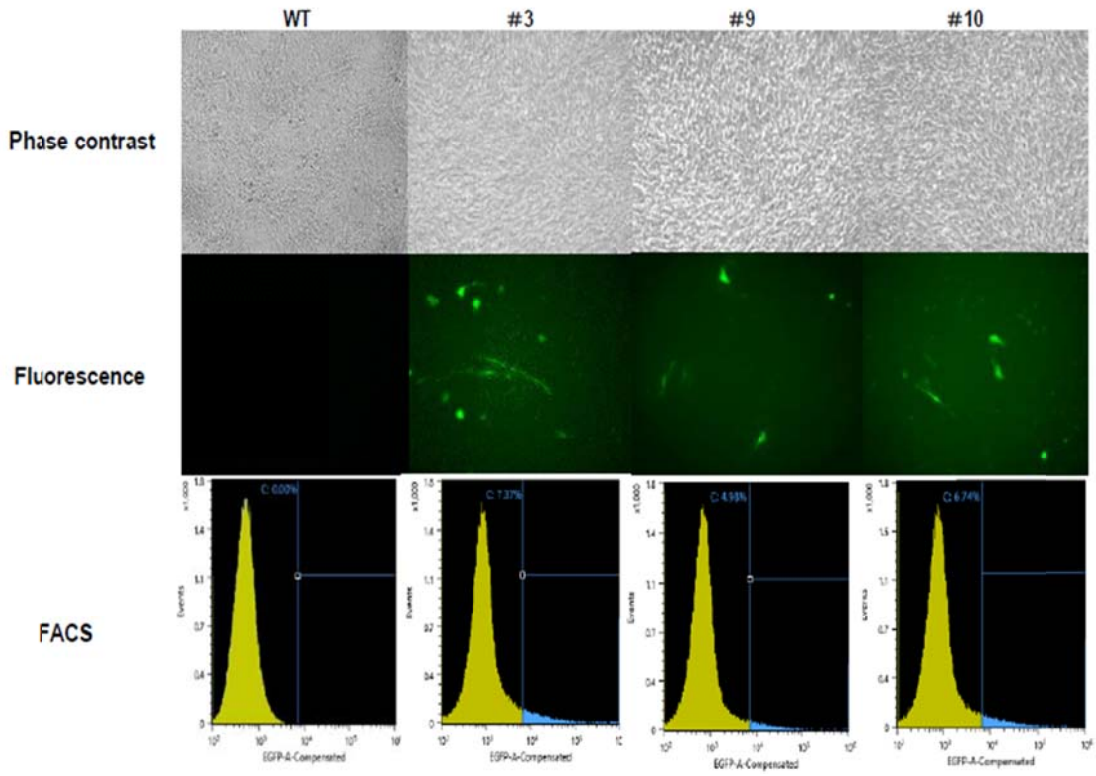


Figure. 5-3 Schematic of fluorescence microscope images and FACS result of *EGFP* expression in clone3#, clone9# and clone10# after transfected dCas9-VPR transactivation system.

5.4.3 Conclusion

By the transfection of dCas9-VPR and gRNA vectors to the knock-in clones without the promoter for *EGFP* (clones no. 3, 9, and 10), *EGFP* expression was successfully induced by endogenous *OVA* promoter in the cells. However, the efficiency of *EGFP*-expressing cells was less than 10%. This is possibly due to the low transfection efficiency upon using the large dCas9-VPR vector (over 13 kb) and multiple vectors for gRNA (total of six plasmids). Nevertheless, a detectable level of *OVA* was expressed in the total cells by the transfection with the vectors.

Chapter 6

Conclusion

Recently, Oishi et al. generated transgenic chickens producing human interferon, in which transgenes were knocked in into the *OVA* locus of chicken primordial germ cells (PGCs) using the CRISPR/Cas9 system and the knock-in PGCs were used for generating transgenic chickens [184]. Transgenes were introduced into exon 2 of the *OVA* gene and a large amount of target protein was produced in egg white. The HITI-mediated CRISPR/Cas9 system could be applied for generating transgenic chickens by using knock-in PGCs or by direct injection into blastodermal-stage embryos.

The premise of high expression of exogenous gene in oviduct is that the selected expression regulation sequence can not only guide the effective expression of exogenous gene in oviduct epithelial cells, but also the expression product can be effectively secreted into egg white. Ovalbumin protein over 50% in egg white, its gene and regulatory region sequence have been studied in detail, and widely used in the construction of oviduct specific expression vector. The important elements that regulate the expression of ovalbumin gene include transcription regulatory elements, elements that affect chromosomal aberrations and open sequences, and elements related to post transcriptional translation, which are induced by steroid hormones such as estrogen, insulin and glucocorticoid. There are two important cis acting elements controlling gene transcription in the 5 '-regulatory region of ovalbumin gene, including two hormone induced DNase 1 hypersensitive sites, in which site I (NRE) is a group of negative regulatory elements located in the –308 ~ –88 region, including the tissue-specific expression regulatory elements of ovalbumin gene; Site II (SDRE) is a sterol hormone dependent regulatory element distributed in the –892 ~ –780 region. NRE has a dual regulatory effect. In the absence of steroids, NRE can inhibit the expression of ovalbumin gene, while in the moderate hormone induction, NRE

can activate gene transcription through the synergistic effect with SDRE.

Chapter 1 introduces general information on the recombination protein expression system, including the most common expression systems of *E. coli*, *Pichia pastoris*, baculovirus/insect cell, mammalian cells, and transgenic chicken oviduct bioreactor. Additionally, the difficulty in the generation of transgenic chicken is introduced. Furthermore, the mechanisms of gene editing methods are also discussed.

In Chapter 2, the techniques and mechanisms related to the study in this thesis are reviewed.

In Chapter 3, the selection of high efficiency gRNAs for the ovalbumin and lysozyme genes in the CEF cells is discussed. The HITI pathway was used to evaluate the knock-in efficiency of an exogenous gene. We obtained gRNA sequences in chicken ovalbumin, which are 31.1%. We utilized the gRNA locus in ovalbumin gene to knock-in *EGFP* cassette into DF-1 cells via homology-independent targeted integration. In total, 20 clones in 1×10^5 cells contained the knock-in gene, which was confirmed by genome PCR test and FACS.

In Chapter 4, dCas9 system with 5 gRNA sequences around the ovalbumin gene promoter (TATA box) were co-cultured into CEF and DF-1 cells to initiate ovalbumin protein expression. RT-PCR, qRT-PCR, and WB analyses revealed that the dCas9 system could activate ovalbumin gene expression with high efficiency.

In Chapter 5, An *EGFP* gene expression cassette was introduced into the *OVA* locus of chicken DF-1 using the CRISPR/Cas9 system via homology-independent targeted integration. *EGFP* expression was successfully induced in the knock-in cells by activating the endogenous *OVA* promoter using the dCas9-VPR transactivation system. The combination of gRNAs designed around the *OVA* TATA box was important to induce the endogenous *OVA* gene expression with high efficiency.

In Chapter 6, the contents of this study are summarized.

In our research we utilized CRISPR/Cas9 system to fully take advantage of 5'-regulatory region. First we designed gRNA sequence in *OVA* promoter area and effectively generated targeted knock-in cells into the *OVA* locus using the

HITI-mediated CRISPR/Cas9 system. Second we successfully activated the endogenous *OVA* gene using the dCas9-VPR transactivation system in chicken non-oviduct cells (DF-1 and CEF cells). The combination of gRNAs designed around the *OVA* TATA box in 5'- regulatory region was important to induce the expression of the endogenous *OVA* gene with high efficiency. Then we confirm endogenous *OVA* promoter was able to initiate exogenous gene expression by using the dCas9-VPR transactivation system. These results are useful for studies on creating transgenic chicken bioreactors and activating tissue-specific promoters.

References

[1] K.T. Cheng, C.L. Wu, B.S. Yip, H.Y. Yu, H.T. Cheng, Y.H. Chih, J.W. Cheng, High Level Expression and Purification of the Clinically Active Antimicrobial Peptide P-113 in Escherichia coli, Molecules 23(4) (2018).

[2] H.P. Sorensen, K.K. Mortensen, Advanced genetic strategies for recombinant protein expression in Escherichia coli, J Biotechnol 115(2) (2005) 113-28.

[3] J. Ou, T. Yamada, K. Nagahisa, T. Hirasawa, C. Furusawa, T. Yomo, H. Shimizu, Dynamic change in promoter activation during lysine biosynthesis in Escherichia coli cells, Mol Biosyst 4(2) (2008) 128-34.

[4] F. Baneyx, Recombinant protein expression in Escherichia coli, Curr Opin Biotechnol 10(5) (1999) 411-21.

[5] T. Han, H. Ming, L. Deng, H. Zhu, Z. Liu, J. Zhang, Y. Song, A novel expression vector for the improved solubility of recombinant scorpion venom in Escherichia coli, Biochem Biophys Res Commun 482(1) (2017) 120-125.

[6] J. Sambrook, D.W. Russell, Expression of Cloned Genes in E. coli Using the Bacteriophage lambda pL Promoter, CSH Protoc 2006(1) (2006).

[7] M. Luo, M. Zhao, C. Cagliero, H. Jiang, Y. Xie, J. Zhu, H. Yang, M. Zhang, Y. Zheng, Y. Yuan, Z. Du, H. Lu, A general platform for efficient extracellular expression and purification of Fab from Escherichia coli, Appl Microbiol Biotechnol 103(8) (2019) 3341-3353.

[8] S. Banerjee, S.S. Salunkhe, A.D. Apte-Deshpande, N.S. Mandi, G. Mandal, S. Padmanabhan, Over-expression of proteins using a modified pBAD24 vector in E. coli expression system, Biotechnol Lett 31(7) (2009) 1031-6.

[9] Adivitiya, V.K. Dagar, N. Devi, Y.P. Khasa, High level production of active streptokinase in Pichia pastoris fed-batch culture, Int J Biol Macromol 83 (2016) 50-60.

[10] J.C. Goodrick, M. Xu, R. Finnegan, B.M. Schilling, S. Schiavi, H. Hoppe, N.C. Wan, High-level expression and stabilization of recombinant human chitinase

produced in a continuous constitutive Pichia pastoris expression system, Biotechnol Bioeng 74(6) (2001) 492-7.

[11] T. Boehm, S. Pirie-Shepherd, L.B. Trinh, J. Shiloach, J. Folkman, Disruption of the KEX1 gene in Pichia pastoris allows expression of full-length murine and human endostatin, Yeast 15(7) (1999) 563-72.

[12] K. Schroer, K. Peter Luef, F. Stefan Hartner, A. Glieder, B. Pscheidt, Engineering the Pichia pastoris methanol oxidation pathway for improved NADH regeneration during whole-cell biotransformation, Metab Eng 12(1) (2010) 8-17.

[13] A. Bronikowski, P.L. Hagedoorn, K. Koschorreck, V.B. Urlacher, Expression of a new laccase from Moniliophthora roreri at high levels in Pichia pastoris and its potential application in micropollutant degradation, AMB Express 7(1) (2017) 73.

[14] R.R. Rowland, B. Robinson, J. Stefanick, T.S. Kim, L. Guanghua, S.R. Lawson, D.A. Benfield, Inhibition of porcine reproductive and respiratory syndrome virus by interferon-gamma and recovery of virus replication with 2-aminopurine, Arch Virol 146(3) (2001) 539-55.

[15] G.E. Smith, M.D. Summers, M.J. Fraser, Production of human beta interferon in insect cells infected with a baculovirus expression vector, Mol Cell Biol 3(12) (1983) 2156-65.

[16] Q. Yao, P. Qian, Y. Cao, Y. He, Y. Si, Z. Xu, H. Chen, Synergistic inhibition of pseudorabies virus replication by porcine alpha/beta interferon and gamma interferon in vitro, Eur Cytokine Netw 18(2) (2007) 71-7.

[17] M.D. Ayres, S.C. Howard, J. Kuzio, M. Lopez-Ferber, R.D. Possee, The complete DNA sequence of Autographa californica nuclear polyhedrosis virus, Virology 202(2) (1994) 586-605.

[18] C.H. Ahrens, R.L. Russell, C.J. Funk, J.T. Evans, S.H. Harwood, G.F. Rohrmann, The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome, Virology 229(2) (1997) 381-99.

[19] S. Gomi, K. Majima, S. Maeda, Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus, J Gen Virol 80 (Pt 5) (1999) 1323-37.

[20] J. Kuzio, M.N. Pearson, S.H. Harwood, C.J. Funk, J.T. Evans, J.M. Slavicek, G.F. Rohrmann, Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar, Virology 253(1) (1999) 17-34.

[21] T. Luque, R. Finch, N. Crook, D.R. O'Reilly, D. Winstanley, The complete sequence of the Cydia pomonella granulovirus genome, J Gen Virol 82(Pt 10) (2001) 2531-47.

[22] L.G. Willis, R. Seipp, T.M. Stewart, M.A. Erlandson, D.A. Theilmann, Sequence analysis of the complete genome of Trichoplusia ni single nucleopolyhedrovirus and the identification of a baculoviral photolyase gene, Virology 338(2) (2005) 209-26.

[23] J.W. Todd, A.L. Passarelli, A. Lu, L.K. Miller, Factors regulating baculovirus late and very late gene expression in transient-expression assays, J Virol 70(4) (1996) 2307-17.

[24] D.A. Theilmann, J.K. Chantler, S. Stweart, H.T. Flipsen, J.M. Vlak, N.E. Crook, Characterization of a highly conserved baculovirus structural protein that is specific for occlusion-derived virions, Virology 218(1) (1996) 148-58.

[25] Y. Matsuura, R.D. Possee, H.A. Overton, D.H. Bishop, Baculovirus expression vectors: the requirements for high level expression of proteins, including glycoproteins, J Gen Virol 68 (Pt 5) (1987) 1233-50.

[26] G.W. Blissard, Baculovirus--insect cell interactions, Cytotechnology 20(1-3) (1996) 73-93.

[27] J.T. Flipsen, J.W. Martens, M.M. van Oers, J.M. Vlak, J.W. van Lent, Passage of Autographa californica nuclear polyhedrosis virus through the midgut epithelium of Spodoptera exigua larvae, Virology 208(1) (1995) 328-35.

[28] J.R. Birch, Y. Onakunle, Biopharmaceutical proteins: opportunities and challenges, Methods Mol Biol 308 (2005) 1-16.

[29] F.L. Graham, J. Smiley, W.C. Russell, R. Nairn, Characteristics of a human cell line transformed by DNA from human adenovirus type 5, J Gen Virol 36(1) (1977) 59-74.

[30] G. Shaw, S. Morse, M. Ararat, F.L. Graham, Preferential transformation of

human neuronal cells by human adenoviruses and the origin of HEK 293 cells, FASEB J 16(8) (2002) 869-71.

[31] J. Zhang, X. Liu, A. Bell, R. To, T.N. Baral, A. Azizi, J. Li, B. Cass, Y. Durocher, Transient expression and purification of chimeric heavy chain antibodies, Protein Expr Purif 65(1) (2009) 77-82.

[32] G. Backliwal, M. Hildinger, S. Chenuet, S. Wulhfard, M. De Jesus, F.M. Wurm, Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions, Nucleic Acids Res 36(15) (2008) e96.

[33] A.D. Bandaranayake, S.C. Almo, Recent advances in mammalian protein production, FEBS Lett 588(2) (2014) 253-60.

[34] R.J. Kaufman, P.A. Sharp, Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary dna gene, J Mol Biol 159(4) (1982) 601-21.

[35] H. Pu, L.M. Cashion, P.J. Kretschmer, Z. Liu, Rapid establishment of high-producing cell lines using dicistronic vectors with glutamine synthetase as the selection marker, Mol Biotechnol 10(1) (1998) 17-25.

[36] L.M. Barnes, C.M. Bentley, A.J. Dickson, Advances in animal cell recombinant protein production: GS-NS0 expression system, Cytotechnology 32(2) (2000) 109-23.

[37] W.H. Brondyk, Selecting an appropriate method for expressing a recombinant protein, Methods Enzymol 463 (2009) 131-47.

[38] J.W. Gordon, G.A. Scangos, D.J. Plotkin, J.A. Barbosa, F.H. Ruddle, Genetic transformation of mouse embryos by microinjection of purified DNA, Proc Natl Acad Sci U S A 77(12) (1980) 7380-4.

[39] J.W. Gordon, F.H. Ruddle, Integration and stable germ line transmission of genes injected into mouse pronuclei, Science 214(4526) (1981) 1244-6.

[40] G.H. Swift, R.E. Hammer, R.J. MacDonald, R.L. Brinster, Tissue-specific expression of the rat pancreatic elastase I gene in transgenic mice, Cell 38(3) (1984) 639-46.

[41] D. Hanahan, Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes, Nature 315(6015) (1985) 115-22.

[42] P.A. Overbeek, A.B. Chepelinsky, J.S. Khillan, J. Piatigorsky, H. Westphal, Lens-specific expression and developmental regulation of the bacterial chloramphenicol acetyltransferase gene driven by the murine alpha A-crystallin promoter in transgenic mice, Proc Natl Acad Sci U S A 82(23) (1985) 7815-9.

[43] A.C. Andres, C.A. Schonenberger, B. Groner, L. Hennighausen, M. LeMeur, P. Gerlinger, Ha-ras oncogene expression directed by a milk protein gene promoter: tissue specificity, hormonal regulation, and tumor induction in transgenic mice, Proc Natl Acad Sci U S A 84(5) (1987) 1299-303.

[44] E.M. Bayna, J.M. Rosen, Tissue-specific, high level expression of the rat whey acidic protein gene in transgenic mice, Nucleic Acids Res 18(10) (1990) 2977-85.

[45] J. Denman, M. Hayes, C. O'Day, T. Edmunds, C. Bartlett, S. Hirani, K.M. Ebert, K. Gordon, J.M. McPherson, Transgenic expression of a variant of human tissue-type plasminogen activator in goat milk: purification and characterization of the recombinant enzyme, Biotechnology (N Y) 9(9) (1991) 839-43.

[46] V.G. Pursel, C.A. Pinkert, K.F. Miller, D.J. Bolt, R.G. Campbell, R.D. Palmiter, R.L. Brinster, R.E. Hammer, Genetic engineering of livestock, Science 244(4910) (1989) 1281-8.

[47] G. Wright, A. Carver, D. Cottom, D. Reeves, A. Scott, P. Simons, I. Wilmut, I. Garner, A. Colman, High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep, Biotechnology (N Y) 9(9) (1991) 830-4.

[48] W.H. Velander, J.L. Johnson, R.L. Page, C.G. Russell, A. Subramanian, T.D. Wilkins, F.C. Gwazdauskas, C. Pittius, W.N. Drohan, High-level expression of a heterologous protein in the milk of transgenic swine using the cDNA encoding human protein C, Proc Natl Acad Sci U S A 89(24) (1992) 12003-7.

[49] A.J. Harvey, G. Speksnijder, L.R. Baugh, J.A. Morris, R. Ivarie, Expression of exogenous protein in the egg white of transgenic chickens, Nat Biotechnol 20(4)

(2002) 396-9.

[50] T.S. Raju, J.B. Briggs, S.M. Borge, A.J. Jones, Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics, Glycobiology 10(5) (2000) 477-86.

[51] M.M. Perry, Nuclear events from fertilisation to the early cleavage stages in the domestic fowl (Gallus domesticus), J Anat 150 (1987) 99-109.

[52] D. Waddington, C. Gribbin, R.J. Sterling, H.M. Sang, M.M. Perry, Chronology of events in the first cell cycle of the polyspermic egg of the domestic fowl (Gallus domesticus), Int J Dev Biol 42(4) (1998) 625-8.

[53] P.E. Mozdziak, J.N. Petitte, Status of transgenic chicken models for developmental biology, Dev Dyn 229(3) (2004) 414-21.

[54] M. Kamihira, K. Nishijima, S. Iijima, Transgenic birds for the production of recombinant proteins, Adv Biochem Eng Biotechnol 91 (2004) 171-89.

[55] R. Heilig, R. Muraskowsky, J.L. Mandel, The ovalbumin gene family. The 5' end region of the X and Y genes, J Mol Biol 156(1) (1982) 1-19.

[56] J.S. Kaye, M. Bellard, G. Dretzen, F. Bellard, P. Chambon, A close association between sites of DNase I hypersensitivity and sites of enhanced cleavage by micrococcal nuclease in the 5'-flanking region of the actively transcribed ovalbumin gene, EMBO J 3(5) (1984) 1137-44.

[57] J.S. Kaye, S. Pratt-Kaye, M. Bellard, G. Dretzen, F. Bellard, P. Chambon, Steroid hormone dependence of four DNase I-hypersensitive regions located within the 7000-bp 5'-flanking segment of the ovalbumin gene, EMBO J 5(2) (1986) 277-85.

[58] S.A. Haecker, T. Muramatsu, K.R. Sensenbaugh, M.M. Sanders, Repression of the ovalbumin gene involves multiple negative elements including a ubiquitous transcriptional silencer, Mol Endocrinol 9(9) (1995) 1113-26.

[59] M.M. Sanders, G.S. McKnight, Positive and negative regulatory elements control the steroid-responsive ovalbumin promoter, Biochemistry 27(17) (1988) 6550-7.

[60] S. Kato, L. Tora, J. Yamauchi, S. Masushige, M. Bellard, P. Chambon, A far

upstream estrogen response element of the ovalbumin gene contains several half-palindromic 5'-TGACC-3' motifs acting synergistically, Cell 68(4) (1992) 731-42.

[61] H.M. Park, S.E. Haecker, S.G. Hagen, M.M. Sanders, COUP-TF plays a dual role in the regulation of the ovalbumin gene, Biochemistry 39(29) (2000) 8537-45.

[62] R. Ivarie, Avian transgenesis: progress towards the promise, Trends Biotechnol 21(1) (2003) 14-9.

[63] (!!! INVALID CITATION !!! {}).

[64] R. Heilig, F. Perrin, F. Gannon, J.L. Mandel, P. Chambon, The ovalbumin gene family: structure of the X gene and evolution of duplicated split genes, Cell 20(3) (1980) 625-37.

[65] T. Muramatsu, T. Imai, H.M. Park, H. Watanabe, A. Nakamura, J. Okumura, Gene gun-mediated in vivo analysis of tissue-specific repression of gene transcription driven by the chicken ovalbumin promoter in the liver and oviduct of laying hens, Mol Cell Biochem 185(1-2) (1998) 27-32.

[66] R.A. Kopper, T. Stallcup, G. Hufford, C.D. Liarakos, The 5'-end structure of ovalbumin mRNA in isolated nuclei and polysomes, Nucleic Acids Res 22(21) (1994) 4504-9.

[67] T. Muramatsu, H. Hiramatsu, J. Okumura, Induction of ovalbumin mRNA by ascorbic acid in primary cultures of tubular gland cells of the chicken oviduct, Comp Biochem Physiol B Biochem Mol Biol 112(2) (1995) 209-16.

[68] M.R. Capecchi, Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century, Nat Rev Genet 6(6) (2005) 507-12.

[69] A. Hinnen, J.B. Hicks, G.R. Fink, Transformation of yeast, Proc Natl Acad Sci U S A 75(4) (1978) 1929-33.

[70] K.R. Thomas, K.R. Folger, M.R. Capecchi, High frequency targeting of genes to specific sites in the mammalian genome, Cell 44(3) (1986) 419-28.

[71] M. Bibikova, D. Carroll, D.J. Segal, J.K. Trautman, J. Smith, Y.G. Kim, S. Chandrasegaran, Stimulation of homologous recombination through targeted cleavage

by chimeric nucleases, Mol Cell Biol 21(1) (2001) 289-97.

[72] S. Sarbajna, S.C. West, Holliday junction processing enzymes as guardians of genome stability, Trends Biochem Sci 39(9) (2014) 409-19.

[73] K.K. Chiruvella, Z. Liang, T.E. Wilson, Repair of double-strand breaks by end joining, Cold Spring Harb Perspect Biol 5(5) (2013) a012757.

[74] A.F. Gilles, M. Averof, Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution, Evodevo 5 (2014) 43.

[75] K.R. Thomas, M.R. Capecchi, Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells, Cell 51(3) (1987) 503-12.

[76] T. Sakuma, K. Woltjen, Nuclease-mediated genome editing: At the front-line of functional genomics technology, Dev Growth Differ 56(1) (2014) 2-13.

[77] D. Kwart, D. Paquet, S. Teo, M. Tessier-Lavigne, Precise and efficient scarless genome editing in stem cells using CORRECT, Nat Protoc 12(2) (2017) 329-354.

[78] J.C. Epinat, S. Arnould, P. Chames, P. Rochaix, D. Desfontaines, C. Puzin, A. Patin, A. Zanghellini, F. Paques, E. Lacroix, A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells, Nucleic Acids Res 31(11) (2003) 2952-62.

[79] Y.G. Kim, J. Cha, S. Chandrasegaran, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain, Proc Natl Acad Sci U S A 93(3) (1996) 1156-60.

[80] S.K. Gupta, P. Shukla, Gene editing for cell engineering: trends and applications, Crit Rev Biotechnol 37(5) (2017) 672-684.

[81] A.A. Chugunova, O.A. Dontsova, P.V. Sergiev, Methods of Genome Engineering: a New Era of Molecular Biology, Biochemistry (Mosc) 81(7) (2016) 662-77.

[82] L. Li, L.P. Wu, S. Chandrasegaran, Functional domains in Fok I restriction endonuclease, Proc Natl Acad Sci U S A 89(10) (1992) 4275-9.

[83] Y.G. Kim, S. Chandrasegaran, Chimeric restriction endonuclease, Proc Natl Acad Sci U S A 91(3) (1994) 883-7.

[84] J. Smith, M. Bibikova, F.G. Whitby, A.R. Reddy, S. Chandrasegaran, D. Carroll, Requirements for double-strand cleavage by chimeric restriction enzymes with zinc

finger DNA-recognition domains, Nucleic Acids Res 28(17) (2000) 3361-9.

[85] Y. Shimizu, M.S. Bhakta, D.J. Segal, Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker, Bioorg Med Chem Lett 19(14) (2009) 3970-2.

[86] M. Bibikova, M. Golic, K.G. Golic, D. Carroll, Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases, Genetics 161(3) (2002) 1169-75.

[87] J.E. Foley, M.L. Maeder, J. Pearlberg, J.K. Joung, R.T. Peterson, J.R. Yeh, Targeted mutagenesis in zebrafish using customized zinc-finger nucleases, Nat Protoc 4(12) (2009) 1855-67.

[88] J. Hauschild, B. Petersen, Y. Santiago, A.L. Queisser, J.W. Carnwath, A. Lucas-Hahn, L. Zhang, X. Meng, P.D. Gregory, R. Schwinzer, G.J. Cost, H. Niemann, Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases, Proc Natl Acad Sci U S A 108(29) (2011) 12013-7.

[89] M. Meyer, M.H. de Angelis, W. Wurst, R. Kuhn, Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases, Proc Natl Acad Sci U S A 107(34) (2010) 15022-6.

[90] Y. Takasu, I. Kobayashi, K. Beumer, K. Uchino, H. Sezutsu, S. Sajwan, D. Carroll, T. Tamura, M. Zurovec, Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection, Insect Biochem Mol Biol 40(10) (2010) 759-65.

[91] J. Zou, M.L. Maeder, P. Mali, S.M. Pruett-Miller, S. Thibodeau-Beganny, B.K. Chou, G. Chen, Z. Ye, I.H. Park, G.Q. Daley, M.H. Porteus, J.K. Joung, L. Cheng, Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells, Cell Stem Cell 5(1) (2009) 97-110.

[92] J. Boch, U. Bonas, Xanthomonas AvrBs3 family-type III effectors: discovery and function, Annu Rev Phytopathol 48 (2010) 419-36.

[93] A.J. Bogdanove, S. Schornack, T. Lahaye, TAL effectors: finding plant genes for disease and defense, Curr Opin Plant Biol 13(4) (2010) 394-401.

[94] R. Morbitzer, P. Romer, J. Boch, T. Lahaye, Regulation of selected genome loci

using de novo-engineered transcription activator-like effector (TALE)-type transcription factors, Proc Natl Acad Sci U S A 107(50) (2010) 21617-22.

[95] J.P. Guilinger, D.B. Thompson, D.R. Liu, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification, Nat Biotechnol 32(6) (2014) 577-582.

[96] C. Ferguson, M. McKay, R.A. Harris, G.E. Homanics, Toll-like receptor 4 (Tlr4) knockout rats produced by transcriptional activator-like effector nuclease (TALEN)-mediated gene inactivation, Alcohol 47(8) (2013) 595-9.

[97] A. Forsyth, T. Weeks, C. Richael, H. Duan, Transcription Activator-Like Effector Nucleases (TALEN)-Mediated Targeted DNA Insertion in Potato Plants, Front Plant Sci 7 (2016) 1572.

[98] N. Sun, J. Liang, Z. Abil, H. Zhao, Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease, Mol Biosyst 8(4) (2012) 1255-63.

[99] T. Watanabe, S. Noji, T. Mito, Gene knockout by targeted mutagenesis in a hemimetabolous insect, the two-spotted cricket Gryllus bimaculatus, using TALENs, Methods 69(1) (2014) 17-21.

[100] Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, A. Nakata, Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product, J Bacteriol 169(12) (1987) 5429-33.

[101] P.W. Hermans, D. van Soolingen, E.M. Bik, P.E. de Haas, J.W. Dale, J.D. van Embden, Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains, Infect Immun 59(8) (1991) 2695-705.

[102] F.J. Mojica, G. Juez, F. Rodriguez-Valera, Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites, Mol Microbiol 9(3) (1993) 613-21.

[103] A. Nakata, M. Amemura, K. Makino, Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome, J Bacteriol 171(6)

(1989) 3553-6.

[104] F.J. Mojica, C. Diez-Villasenor, E. Soria, G. Juez, Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria, Mol Microbiol 36(1) (2000) 244-6.

[105] R. Jansen, J.D. Embden, W. Gaastra, L.M. Schouls, Identification of genes that are associated with DNA repeats in prokaryotes, Mol Microbiol 43(6) (2002) 1565-75.

[106] A. Bolotin, B. Quinquis, A. Sorokin, S.D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology 151(Pt 8) (2005) 2551-61.

[107] C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, Microbiology 151(Pt 3) (2005) 653-63.

[108] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes, Science 315(5819) (2007) 1709-12.

[109] J.E. Garneau, M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A.H. Magadan, S. Moineau, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA, Nature 468(7320) (2010) 67-71.

[110] E.R. Westra, D.C. Swarts, R.H. Staals, M.M. Jore, S.J. Brouns, J. van der Oost, The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity, Annu Rev Genet 46 (2012) 311-39.

[111] Y. Agari, K. Sakamoto, M. Tamakoshi, T. Oshima, S. Kuramitsu, A. Shinkai, Transcription profile of Thermus thermophilus CRISPR systems after phage infection, J Mol Biol 395(2) (2010) 270-81.

[112] K. Pougach, E. Semenova, E. Bogdanova, K.A. Datsenko, M. Djordjevic, B.L. Wanner, K. Severinov, Transcription, processing and function of CRISPR cassettes in Escherichia coli, Mol Microbiol 77(6) (2010) 1367-79.

[113] V. Kunin, R. Sorek, P. Hugenholtz, Evolutionary conservation of sequence and secondary structures in CRISPR repeats, Genome Biol 8(4) (2007) R61.

[114] A. Kuzminov, Single-strand interruptions in replicating chromosomes cause double-strand breaks, Proc Natl Acad Sci U S A 98(15) (2001) 8241-6.

[115] I. Yosef, M.G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli, Nucleic Acids Res 40(12) (2012) 5569-76.

[116] K.S. Makarova, L. Aravind, N.V. Grishin, I.B. Rogozin, E.V. Koonin, A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis, Nucleic Acids Res 30(2) (2002) 482-96.

[117] D.H. Haft, J. Selengut, E.F. Mongodin, K.E. Nelson, A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes, PLoS Comput Biol 1(6) (2005) e60.

[118] K.S. Makarova, D.H. Haft, R. Barrangou, S.J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F.J. Mojica, Y.I. Wolf, A.F. Yakunin, J. van der Oost, E.V. Koonin, Evolution and classification of the CRISPR-Cas systems, Nat Rev Microbiol 9(6) (2011) 467-77.

[119] N. Beloglazova, P. Petit, R. Flick, G. Brown, A. Savchenko, A.F. Yakunin, Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference, EMBO J 30(22) (2011) 4616-27.

[120] R. Sapranauskas, G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath, V. Siksnys, The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli, Nucleic Acids Res 39(21) (2011) 9275-82.

[121] E. Deltcheva, K. Chylinski, C.M. Sharma, K. Gonzales, Y. Chao, Z.A. Pirzada, M.R. Eckert, J. Vogel, E. Charpentier, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III, Nature 471(7340) (2011) 602-7.

[122] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, Science 337(6096) (2012) 816-21.

[123] L.A. Marraffini, E.J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA, Science 322(5909) (2008) 1843-5.

[124] L.A. Marraffini, E.J. Sontheimer, CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea, Nat Rev Genet 11(3) (2010) 181-90.

[125] C.R. Hale, P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, R.M. Terns, M.P. Terns, RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex, Cell 139(5) (2009) 945-56.

[126] K.P. Kim, E.V. Mirkin, So similar yet so different: The two ends of a double strand break, Mutat Res 809 (2018) 70-80.

[127] D. Arosio, S. Cui, C. Ortega, M. Chovanec, S. Di Marco, G. Baldini, A. Falaschi, A. Vindigni, Studies on the mode of Ku interaction with DNA, J Biol Chem 277(12) (2002) 9741-8.

[128] K. Meek, S. Gupta, D.A. Ramsden, S.P. Lees-Miller, The DNA-dependent protein kinase: the director at the end, Immunol Rev 200 (2004) 132-41.

[129] E. Riballo, M. Kuhne, N. Rief, A. Doherty, G.C. Smith, M.J. Recio, C. Reis, K. Dahm, A. Fricke, A. Krempler, A.R. Parker, S.P. Jackson, A. Gennery, P.A. Jeggo, M. Lobrich, A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci, Mol Cell 16(5) (2004) 715-24.

[130] T. Mimori, M. Akizuki, H. Yamagata, S. Inada, S. Yoshida, M. Homma, Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap, J Clin Invest 68(3) (1981) 611-20.

[131] M.J. Difilippantonio, J. Zhu, H.T. Chen, E. Meffre, M.C. Nussenzweig, E.E. Max, T. Ried, A. Nussenzweig, DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation, Nature 404(6777) (2000) 510-4.

[132] Y. Gu, K.J. Seidl, G.A. Rathbun, C. Zhu, J.P. Manis, N. van der Stoep, L. Davidson, H.L. Cheng, J.M. Sekiguchi, K. Frank, P. Stanhope-Baker, M.S. Schlissel, D.B. Roth, F.W. Alt, Growth retardation and leaky SCID phenotype of Ku70-deficient mice, Immunity 7(5) (1997) 653-65.

[133] G.C. Smith, S.P. Jackson, The DNA-dependent protein kinase, Genes Dev 13(8) (1999) 916-34.

[134] S.Y. Shieh, M. Ikeda, Y. Taya, C. Prives, DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2, Cell 91(3) (1997) 325-34.

[135] M. Martin, A. Genesca, L. Latre, I. Jaco, G.E. Taccioli, J. Egozcue, M.A. Blasco, G. Iliakis, L. Tusell, Postreplicative joining of DNA double-strand breaks causes genomic instability in DNA-PKcs-deficient mouse embryonic fibroblasts, Cancer Res 65(22) (2005) 10223-32.

[136] Z. Li, T. Otevrel, Y. Gao, H.L. Cheng, B. Seed, T.D. Stamato, G.E. Taccioli, F.W. Alt, The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination, Cell 83(7) (1995) 1079-89.

[137] S.N. Andres, A. Vergnes, D. Ristic, C. Wyman, M. Modesti, M. Junop, A human XRCC4-XLF complex bridges DNA, Nucleic Acids Res 40(4) (2012) 1868-78.

[138] U. Grawunder, M. Wilm, X. Wu, P. Kulesza, T.E. Wilson, M. Mann, M.R. Lieber, Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells, Nature 388(6641) (1997) 492-5.

[139] D. Moshous, I. Callebaut, R. de Chasseval, B. Corneo, M. Cavazzana-Calvo, F. Le Deist, I. Tezcan, O. Sanal, Y. Bertrand, N. Philippe, A. Fischer, J.P. de Villartay, Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency, Cell 105(2) (2001) 177-86.

[140] A. Decottignies, Alternative end-joining mechanisms: a historical perspective, Front Genet 4 (2013) 48.

[141] M.R. Lieber, T.E. Wilson, SnapShot: Nonhomologous DNA end joining (NHEJ), Cell 142(3) (2010) 496-496 e1.

[142] L. Deriano, D.B. Roth, Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage, Annu Rev Genet 47 (2013) 433-55.

[143] C.T. Yan, C. Boboila, E.K. Souza, S. Franco, T.R. Hickernell, M. Murphy, S. Gumaste, M. Geyer, A.A. Zarrin, J.P. Manis, K. Rajewsky, F.W. Alt, IgH class switching and translocations use a robust non-classical end-joining pathway, Nature

449(7161) (2007) 478-82.

[144] B. Corneo, R.L. Wendland, L. Deriano, X. Cui, I.A. Klein, S.Y. Wong, S. Arnal, A.J. Holub, G.R. Weller, B.A. Pancake, S. Shah, V.L. Brandt, K. Meek, D.B. Roth, Rag mutations reveal robust alternative end joining, Nature 449(7161) (2007) 483-6.

[145] S. Ramakrishna, S.W. Cho, S. Kim, M. Song, R. Gopalappa, J.S. Kim, H. Kim, Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations, Nat Commun 5 (2014) 3378.

[146] P. Perez-Pinera, D.D. Kocak, C.M. Vockley, A.F. Adler, A.M. Kabadi, L.R. Polstein, P.I. Thakore, K.A. Glass, D.G. Ousterout, K.W. Leong, F. Guilak, G.E. Crawford, T.E. Reddy, C.A. Gersbach, RNA-guided gene activation by CRISPR-Cas9-based transcription factors, Nat Methods 10(10) (2013) 973-6.

[147] M.H. Larson, L.A. Gilbert, X. Wang, W.A. Lim, J.S. Weissman, L.S. Qi, CRISPR interference (CRISPRi) for sequence-specific control of gene expression, Nat Protoc 8(11) (2013) 2180-96.

[148] L.A. Gilbert, M.H. Larson, L. Morsut, Z. Liu, G.A. Brar, S.E. Torres, N. Stern-Ginossar, O. Brandman, E.H. Whitehead, J.A. Doudna, W.A. Lim, J.S. Weissman, L.S. Qi, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes, Cell 154(2) (2013) 442-51.

[149] L.S. Qi, M.H. Larson, L.A. Gilbert, J.A. Doudna, J.S. Weissman, A.P. Arkin, W.A. Lim, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, Cell 152(5) (2013) 1173-83.

[150] L.A. Gilbert, M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, C. Guimaraes, B. Panning, H.L. Ploegh, M.C. Bassik, L.S. Qi, M. Kampmann, J.S. Weissman, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation, Cell 159(3) (2014) 647-61.

[151] A. Plessis, A. Perrin, J.E. Haber, B. Dujon, Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus, Genetics 130(3) (1992) 451-60.

[152] S. Lin, B. Ewen-Campen, X. Ni, B.E. Housden, N. Perrimon, In Vivo

Transcriptional Activation Using CRISPR/Cas9 in Drosophila, Genetics 201(2) (2015) 433-42.

[153] Y. Zhang, C. Yin, T. Zhang, F. Li, W. Yang, R. Kaminski, P.R. Fagan, R. Putatunda, W.B. Young, K. Khalili, W. Hu, CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs, Sci Rep 5 (2015) 16277.

[154] M.L. Maeder, S.J. Linder, V.M. Cascio, Y. Fu, Q.H. Ho, J.K. Joung, CRISPR RNA-guided activation of endogenous human genes, Nat Methods 10(10) (2013) 977-9.

[155] M.E. Tanenbaum, L.A. Gilbert, L.S. Qi, J.S. Weissman, R.D. Vale, A protein-tagging system for signal amplification in gene expression and fluorescence imaging, Cell 159(3) (2014) 635-46.

[156] A. Chavez, J. Scheiman, S. Vora, B.W. Pruitt, M. Tuttle, P.R.I. E, S. Lin, S. Kiani, C.D. Guzman, D.J. Wiegand, D. Ter-Ovanesyan, J.L. Braff, N. Davidsohn, B.E. Housden, N. Perrimon, R. Weiss, J. Aach, J.J. Collins, G.M. Church, Highly efficient Cas9-mediated transcriptional programming, Nat Methods 12(4) (2015) 326-8.

[157] P. Mali, J. Aach, P.B. Stranges, K.M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, G.M. Church, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering, Nat Biotechnol 31(9) (2013) 833-8.

[158] R.D. Palmiter, E.P. Sandgren, M.R. Avarbock, D.D. Allen, R.L. Brinster, Heterologous introns can enhance expression of transgenes in mice, Proc Natl Acad Sci U S A 88(2) (1991) 478-82.

[159] B. Gao, H.C. Sun, C.Y. Song, Z.Y. Wang, Q. Chen, H.Q. Song, Transfection and expression of exogenous gene in laying hens oviduct in vitro and in vivo, J Zhejiang Univ Sci B 6(2) (2005) 137-41.

[160] H. Nishimasu, X. Shi, S. Ishiguro, L. Gao, S. Hirano, S. Okazaki, T. Noda, O.O. Abudayyeh, J.S. Gootenberg, H. Mori, S. Oura, B. Holmes, M. Tanaka, M. Seki, H. Hirano, H. Aburatani, R. Ishitani, M. Ikawa, N. Yachie, F. Zhang, O. Nureki,

Engineered CRISPR-Cas9 nuclease with expanded targeting space, Science 361(6408) (2018) 1259-1262.

[161] L. Cong, F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems, Science 339(6121) (2013) 819-23.

[162] Y. Naito, K. Hino, H. Bono, K. Ui-Tei, CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites, Bioinformatics 31(7) (2015) 1120-3.

[163] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system, Nat Protoc 8(11) (2013) 2281-2308.

[164] C.M. Hammond, C.B. Stromme, H. Huang, D.J. Patel, A. Groth, Histone chaperone networks shaping chromatin function, Nat Rev Mol Cell Biol 18(3) (2017) 141-158.

[165] K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution, Nature 389(6648) (1997) 251-60.

[166] K.W. Trotter, T.K. Archer, Assaying chromatin structure and remodeling by restriction enzyme accessibility, Methods Mol Biol 833 (2012) 89-102.

[167] Y. Rodriguez, J.M. Hinz, M.J. Smerdon, Accessing DNA damage in chromatin: Preparing the chromatin landscape for base excision repair, DNA Repair (Amst) 32 (2015) 113-9.

[168] S.H. Sternberg, S. Redding, M. Jinek, E.C. Greene, J.A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9, Nature 507(7490) (2014) 62-7.

[169] C. Anders, O. Niewoehner, A. Duerst, M. Jinek, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, Nature 513(7519) (2014) 569-73.

[170] M. Jinek, F. Jiang, D.W. Taylor, S.H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A.T. Iavarone, E. Charpentier, E. Nogales, J.A.

Doudna, Structures of Cas9 endonucleases reveal RNA-mediated conformational activation, Science 343(6176) (2014) 1247997.

[171] H. Nishimasu, F.A. Ran, P.D. Hsu, S. Konermann, S.I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, O. Nureki, Crystal structure of Cas9 in complex with guide RNA and target DNA, Cell 156(5) (2014) 935-49.

[172] D. Mashiko, Y. Fujihara, Y. Satouh, H. Miyata, A. Isotani, M. Ikawa, Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA, Sci Rep 3 (2013) 3355.

[173] M. Frank-Vaillant, S. Marcand, NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the ligase IV pathway, Genes Dev 15(22) (2001) 3005-12.

[174] K. Suzuki, Y. Tsunekawa, R. Hernandez-Benitez, J. Wu, J. Zhu, E.J. Kim, F. Hatanaka, M. Yamamoto, T. Araoka, Z. Li, M. Kurita, T. Hishida, M. Li, E. Aizawa, S. Guo, S. Chen, A. Goebl, R.D. Soligalla, J. Qu, T. Jiang, X. Fu, M. Jafari, C.R. Esteban, W.T. Berggren, J. Lajara, E. Nunez-Delicado, P. Guillen, J.M. Campistol, F. Matsuzaki, G.H. Liu, P. Magistretti, K. Zhang, E.M. Callaway, K. Zhang, J.C. Belmonte, In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration, Nature 540(7631) (2016) 144-149.

[175] P.O. Kohler, P.M. Grimley, B.W. O'Malley, Protein synthesis: differential stimulation of cell-specific proteins in epithelial cells of chick oviduct, Science 160(3823) (1968) 86-7.

[176] K. Kasperczyk, A. Bajek, R. Joachimiak, K. Walasik, A. Marszalek, T. Drewa, M. Bednarczyk, In vitro optimization of the Gallus domesticus oviduct epithelial cells culture, Theriogenology 77(9) (2012) 1834-45.

[177] J.D. Sander, J.K. Joung, CRISPR-Cas systems for editing, regulating and targeting genomes, Nat Biotechnol 32(4) (2014) 347-55.

[178] A. Royal, A. Garapin, B. Cami, F. Perrin, J.L. Mandel, M. LeMeur, F. Bregegegre, F. Gannon, J.P. LePennec, P. Chambon, P. Kourilsky, The ovalbumin gene region: common features in the organisation of three genes expressed in chicken

oviduct under hormonal control, Nature 279(5709) (1979) 125-32.
[179] H.M. Park, J. Okumura, T. Muramatsu, Modulation of transcriptional activity of the chicken ovalbumin gene promoter in primary cultures of chicken oviduct cells: effects of putative regulatory elements in the 5'-flanking region, Biochem Mol Biol Int 36(4) (1995) 811-6.
[180] S.G. Lillico, A. Sherman, M.J. McGrew, C.D. Robertson, J. Smith, C. Haslam, P. Barnard, P.A. Radcliffe, K.A. Mitrophanous, E.A. Elliot, H.M. Sang, Oviduct-specific expression of two therapeutic proteins in transgenic hens, Proc Natl Acad Sci U S A 104(6) (2007) 1771-6.
[181] S.J. Byun, S.W. Kim, K.W. Kim, J.S. Kim, I.S. Hwang, H.K. Chung, I.S. Kan, I.S. Jeon, W.K. Chang, S.B. Park, J.G. Yoo, Oviduct-specific enhanced green fluorescent protein expression in transgenic chickens, Biosci Biotechnol Biochem 75(4) (2011) 646-9.
[182] H.A. Kaleri, S.Y. Xu, H.L. Lin, Generation of transgenic chicks using an oviduct-specific expression system, Genet Mol Res 10(4) (2011) 3046-55.
[183] M.S. Kwon, B.C. Koo, D. Kim, Y.H. Nam, X.S. Cui, N.H. Kim, T. Kim, Generation of transgenic chickens expressing the human erythropoietin (hEPO) gene in an oviduct-specific manner: Production of transgenic chicken eggs containing human erythropoietin in egg whites, PLoS One 13(5) (2018) e0194721.
[184] I. Oishi, K. Yoshii, D. Miyahara, T. Tagami, Efficient production of human interferon beta in the white of eggs from ovalbumin gene-targeted hens, Sci Rep 8(1) (2018) 10203.

Acknowledgements

First and foremost, I gratefully acknowledge the help of my supervisor, Prof. Masamichi Kamihira, who has offered me valuable suggestions and whose useful suggestions, incisive comments and constructive criticism have contributed greatly to the completion of this thesis. I do appreciate his patience, encouragement, and professional instructions during my academic studies. I am also grateful to my thesis committee members, Prof. Hiroyuki Ijima and Prof. Noriho Kamiya, for their critical reading of the thesis and their guidance.

I would like to further extend my thanks to Dr. Yoshinori Kawabe and Assoc. Prof. Akira Ito, who have helped me directly and indirectly in my studies, from whose devoted teaching and enlightening lectures I have benefited a lot and overcame difficulties in experiments.

Special thanks also go to all the lab members, past and present. They not only helped me with my research, but also let me know this country better. They are and will always be my friends. I also want to thanks my group members, for their supporting and contribution during my research in Japan.

Last but not the least, my gratitude also extends to my family who have been assisting, supporting and caring for me all of my life. Special thanks should go to my friends who have put considerable time and effort for my study.