

CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats-Cas9 endonuclease), a Nascent and Promising Technology for Genetic Editing in Mammalian Cells

By

Alhan Mehrabi Yazdi

Supervisors: Dr. Ron Shigeta

Dr. Keyvan Sedaghat

This thesis is submitted to the Department of Biology as a partial fulfillment of the BSc program in Biology

Department of Biology
Faculty of Science
Baha'i Institute for Higher Education (BIHE)

Abstract

Genome-editing technologies refer to those systems that have been used to induce changes or modification such as adding or deleting nucleotides on DNA strands. Zinc Finger Nucleases (ZFNs), Transcription Activator-like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats-Cas9 endonuclease (CRISPR-Cas9) are three common genome editing systems that have their own specific properties, advantages, and disadvantages. The main article “A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks” was given by my supervisor Dr Ron Shigeta and by utilizing the Pubmed database, this data is gathered. Parnas and his colleagues showed the efficiency of CRISPR-Cas9 system for finding new regulatory genes in the tumor necrosis factor- α (TNF- α) induction pathway. CRISPR-Cas9 is introduced as the most advanced method, because Cas9 can generate break in both strands of DNA with the help of only one sgRNA. Moreover, one Cas9 can cooperate with multiple sgRNAs and break DNA at multiple sites simultaneously and sgRNAs production just needs to design and synthesis 20 nucleotides which have complementary sequence with our target on DNA. Several innovative achievements have been gained by CRISPR-Cas9 technology, for example, the manipulation and modification of the *Trypanosoma cruzi* (*T. cruzi*) genome, identification of essential genes for West Nile Virus-induced cell death, and its therapeutic approach for individual with Duchenne muscular dystrophy (DMD). In this manuscript, we evaluate how often CRISPR has been successful, while assessing its specificity and sensitivity for finding regulatory genes. Finally, we discuss what factors or strategies could be utilized to improve its efficacy as a gene editing tool.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Figures	v
List of Tables	vi
List of Abbreviations	vii
Acknowledgments	x
Chapter1: Introduction	1
Chapter 2: Methodology	10
Chapter 3: Parnas et al. Research	11
3.1. TNF Induction Pathway and Its Components	11
3.2. Cells and Mice Preparation	14
3.3 Making sgRNA pooled Library	15
3.4. Delivery of CRISPR-Cas9 Components	17
3.5. sgRNA Abundance Determination	18
3.6. Results of Their Work	19
3.7. Some points Regarding the Authors' Research	23
Chapter 4: Advantages of CRISPR-Cas9 System	24
4.1. Cas9 Properties and Potentials	24
4.2. Other Advantages of Cas9	26

4.3. Activating and Repressing Genes by CRISPR-dCas9	27
4.4. CRISPR-Cas9 as a means of DNA visualizing	33
4.5. More Advantages of CRISPR-Cas9 in Comparison with Previous Systems	34
Chapter 5: Other Accomplishments	39
5.1. CRISPR/Cas9 in T. cruzi Genome Screens	39
5.2. Identifying Essential Genes for WNV-induced Cell Death	39
5.3. Therapeutic Approach of CRISPR-Cas9 for Individual with DMD	40
5.4. Treating Tyrosinaemia Type 1 with CRISPR-Cas9.....	41
Chapter 6: Challenges with CRISPR-Cas9	42
6.1. Off-target Effects	42
6.2. PAM Dependence	42
6.3. gRNA Production	42
6.4. Delivery Methods	43
Chapter 7: Conclusion and Discussion	44
References	46

List of Figures

Figure 1: A schematic representation of a zinc-finger nuclease (ZFN) pair.....	2
Figure 2: TAL effector and TALEN structure.....	3
Figure 3: General overview of the two-step cloning strategy for dTALEs assembly.....	5
Figure 4: Schematic structure of the Cas9-sgRNA-DNA ternary complex.....	8
Figure 5: Schematic view of the TLR4 signaling pathway.....	13
Figure 6: Cloning of a miRNA sgRNA library using the Molecular Chipper method	16
Figure 7: Next Generation Sequencing for RNA	19
Figure 8: The validated positive regulators partition into key modules by their effect on protein and RNA expression	22
Figure 9: CRISPR–Cas9-mediated genome editing	26
Figure 10: Genomic Regulatory Programming Using CRISPR and Multi-Domain Scaffolding RNAs.....	32
Figure 11: Imaging of genomic loci using the CRISPR-Cas9 method.....	34

List of Tables

Table 1: Comparison between CRISPR-Cas9 and TALENs	35
Table 2: Table 2: Advantages and Disadvantages of CRISPR-Cas9	44

List of Abbreviations

A-D

Cas protein: CRISPR-associated protein(s)

CD-14: Cluster of Differentiation 14

CMV: Cytomegaloviruse

crRNA: CRISPR RNA

Cys: Cysteine

DMD: Duchenne Muscular Dystrophy

DSB: Double Stranded Breaks

E-G

ER: Endoplasmic Reticulum

FATP: Fatty Acid Transport Protein

GFP: Green Fluorescent Protein

H

HDR: homology-directed repair

His: Histidine

Hpd: hydroxyphenylpyruvate dioxigenase

I-L

IL: Interleukin

LBP: LPS-binding protein

LPS: bacterial Lipopolysaccharide

M-O

Mip-1a: Macrophage inflammatory protein- 1a

MYD: Myeloid Differentiation

NHEJ: Non-homologous End Joining

nt: Nucleotide

P

PAF: RNA polymerase II- associated factor

PAGE: Polyacrylamide Gel Electrophoresis

PAM: Protospacer Adjacent Motif

PCR: Polymerase Chain Reaction

Q- S

SV40: Simian Virus 40

RVD: repeat-variable di-residue

RelA: Rel Associated protein

sgRNA: small guide RNA

T

TNF: Tumor Necrosis Factor

TALENs: Transcription activator-like effector nucleases

tracrRNA: transactivating RNA

TTP: tristetraproline

U-Z

WT: Wild Type

WNV: West Nile Virus

ZFP: Zinc-finger Protein

ZFNs: Zinc-finger nucleases

Acknowledgments

I would like to thank my dear supervisors, Dr. Ron Shigeta who proposed this topic for me, attended me with patience and favor and Dr. Keyvan Sedaghat who was like a light in darkness and gave me hope and inspiration back with his encouragements, guidance and advice for finalizing this project. I also express my gratefulness to my kind parents who are caring and supportive during all my life, all my teachers and instructors who support me scientifically and emotionally, and two of my friends, Dr. Alireza Zahraie and Shayan Moghimy Firooz Abad, who guided me with their experience in writing thesis and articles wholeheartedly.

Chapter 1: Introduction

Recombinant DNA technology was the first evidence of genetic engineering in 1972. In the laboratory of Paul Berg, scientists combined the genes of a bacteriophage and the Simian Vacuolating (SV40) virus with the *Escherichia coli* (*E. coli*) genome. One of the main goals for synthesizing such hybrid DNA was to introduce foreign genes into mammalian cells. Using this approach, they could study the molecular biology of foreign DNA and how it affected the infected mammalian cells. Since then, several studies have been conducted in the field of molecular genetics, which resulted in the development of genome editing technologies (1).

Genome-editing technologies are techniques that are employed to induce changes and modifications in DNA structure and genes (2). These techniques are used to make base-pair substitutions, addition or deletion within DNA strands to determine what the effects are on the phenotype. For example, sometimes these changes within genome led to the correction of mutations or deletion of mutations which cause diseases have had therapeutic effects. Many editing technologies utilize engineered nucleases to achieve this goal (3).

Engineered nuclease induces double-strand breaks in targeted DNA and stimulates the cell to repair DNA by error-prone non-homologous end joining (NHEJ) or homology-directed repair (HDR) (2-5). NHEJ-mediated repair is an example of a DNA editing technology that results in insertion or deletion of small nucleotides at the targeted sites, which can result in turning genes on or off (1, 3,4). In the HDR repair system, specific DNA sequences in the repair template co-deliver with site-specific nucleases to integrate single or multiple transgenes into or near the break site (3, 4). These insertions can affect gene activity (1, 5).

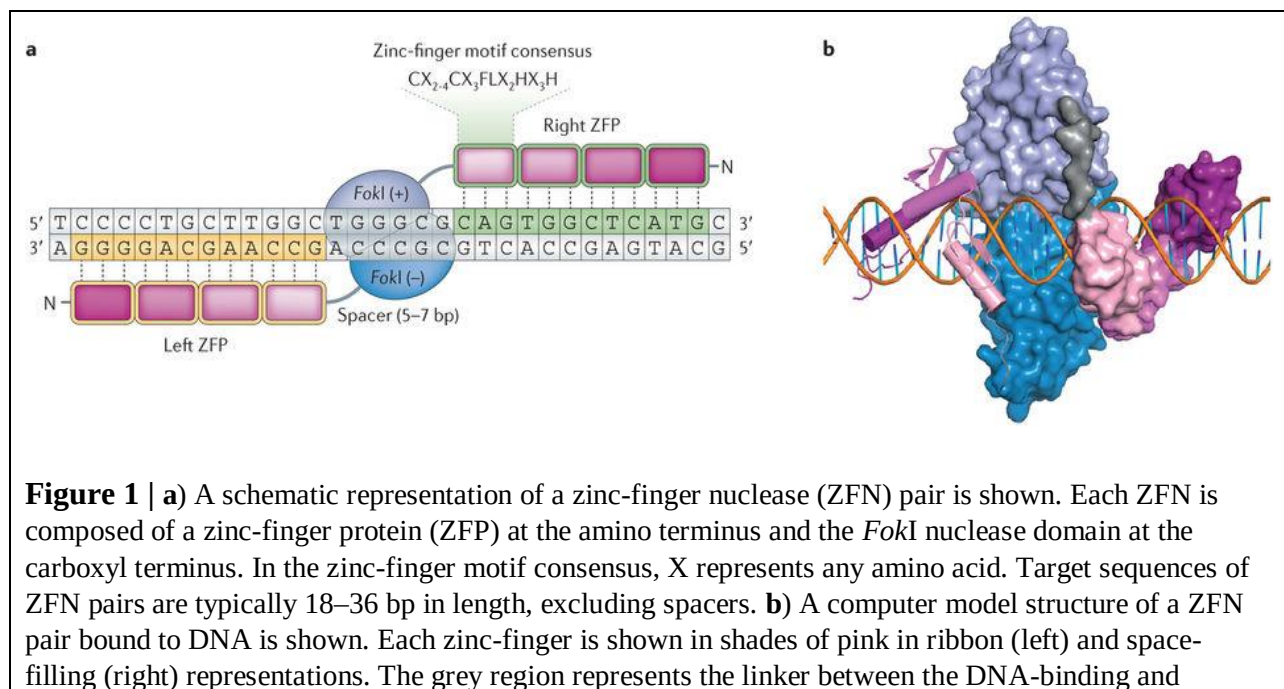
A few methods such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and most recently developed Clustered Regularly Interspaced Short Palindromic Repeats-Cas9 endonuclease (CRISPR-Cas9) system have been employed to edit genes (6, 7). Application of these methods, has made genome modifications possible in a broad range of organisms such as *Drosophila*, rats, mice, zebrafish, and also in diverse human cell types, including primary T lymphocytes, embryonic and induced pluripotent stem cells (7).

Zinc fingers (ZFs) are proteins with modular structure, which can be coupled with an endonuclease domain such as FokI and act as a site-specific nuclease for DNA (1). In the ZFN method, two zinc

finger arrays are engineered (each typically 9-12 bp) that are separated by a spacer sequence (typically 5-7 bp) (4), and each domain of this protein recognizes one nucleotide triplet (1). After binding with target DNA, FokI dimerizes and creates a double-strand break within the spacer region (7) (Fig. 1).

The TALEN system is another efficient method of gene editing. This system has been developed from the *Xanthomonas* bacteria, a genus of crop plants' pathogens. These bacteria secrete transcription activator-like effectors (TALEs) proteins that are capable of binding with the DNA and activate the expression of their target genes (1, 3, 4). Main TALE proteins are 12-27 full repeats of monomers and each monomer is a series of 33-35 amino acid residues (1, 3, 4, 7).

The target specificity is determined by two adjacent residues at positions 12 and 13 that is the repeat-variable di-residue (RVD) in each monomer. Each RVD binds with one nucleotide (1, 3, 4, 7). In most cases, there are four common RVDs (asparagine and isoleucine (NI), histidine and aspartic acid (HD), asparagine and glycine (NG), and two asparagines (NN)) that each is associated with one of the four bases (Fig. 2a). These relationships allow the prediction of TALEN binding site with DNA and the construction of TAL proteins (4). By binding artificial nucleases such as FokI with these chimeric proteins at spacer sequence, double strand breaks are created in target sites (1) (Fig. 2b).



catalytic domains. The *FokI* catalytic domains are shown in blue and purple at the center using space-filling representations (8).

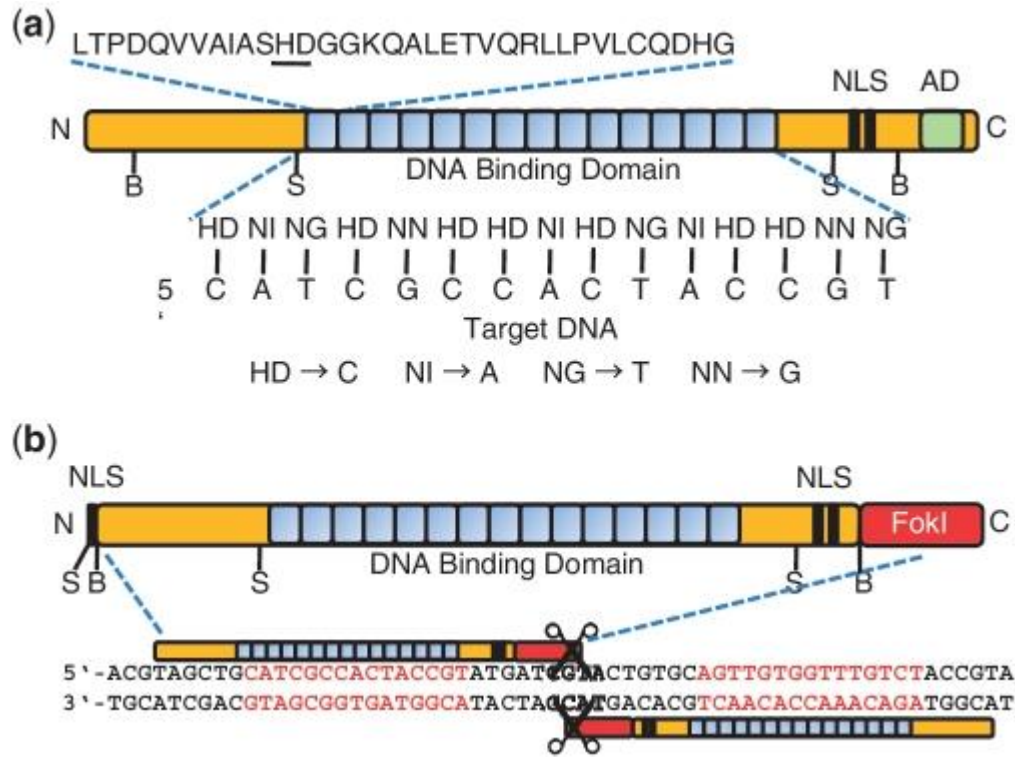


Figure 2 | TAL effector and TALEN structure. (a) Structure of a naturally occurring TAL effector. A consensus repeat sequence is shown with the repeat-variable di-residue (RVD) underlined. The sequence of RVDs determines the target nucleotide sequence. The four most common RVDs are shown with their most frequently associated nucleotide. Some evidence suggests that the less common RVD NK (not displayed) has greater specificity for G than NN does. (b) Structure of a TALEN. Two monomeric TALENs are required to bind the target site to enable FokI to dimerize and cleave DNA. NLS, nuclear localization signal(s); AD, transcriptional activation domain; B, BamHI; S, SphI (9).

Although ZFNs and TALENs are two means for introducing alterations and modifications in genome, there are some limitations of and difficulties with these systems (1, 7). As explained before, it is necessary to construct proteins in both of these systems. Each monomer in zinc finger proteins recognizes only 3 nucleotides. Likewise, in the TALEN system, TALE central binding with DNA contains as many as 30 repeats of a 33 to 35 amino acid sequence motif, each recognizing only a single base pair (4, 7). In addition, it would be necessary to construct two arrays of ZF because FokI must be dimerized for its action as in TALENs. Thus, not only engineering

these amount of amino acids is expensive and time-consuming, but also arranging and putting them in order to pair with the target site it is complex and intricate (5) (Fig. 3).

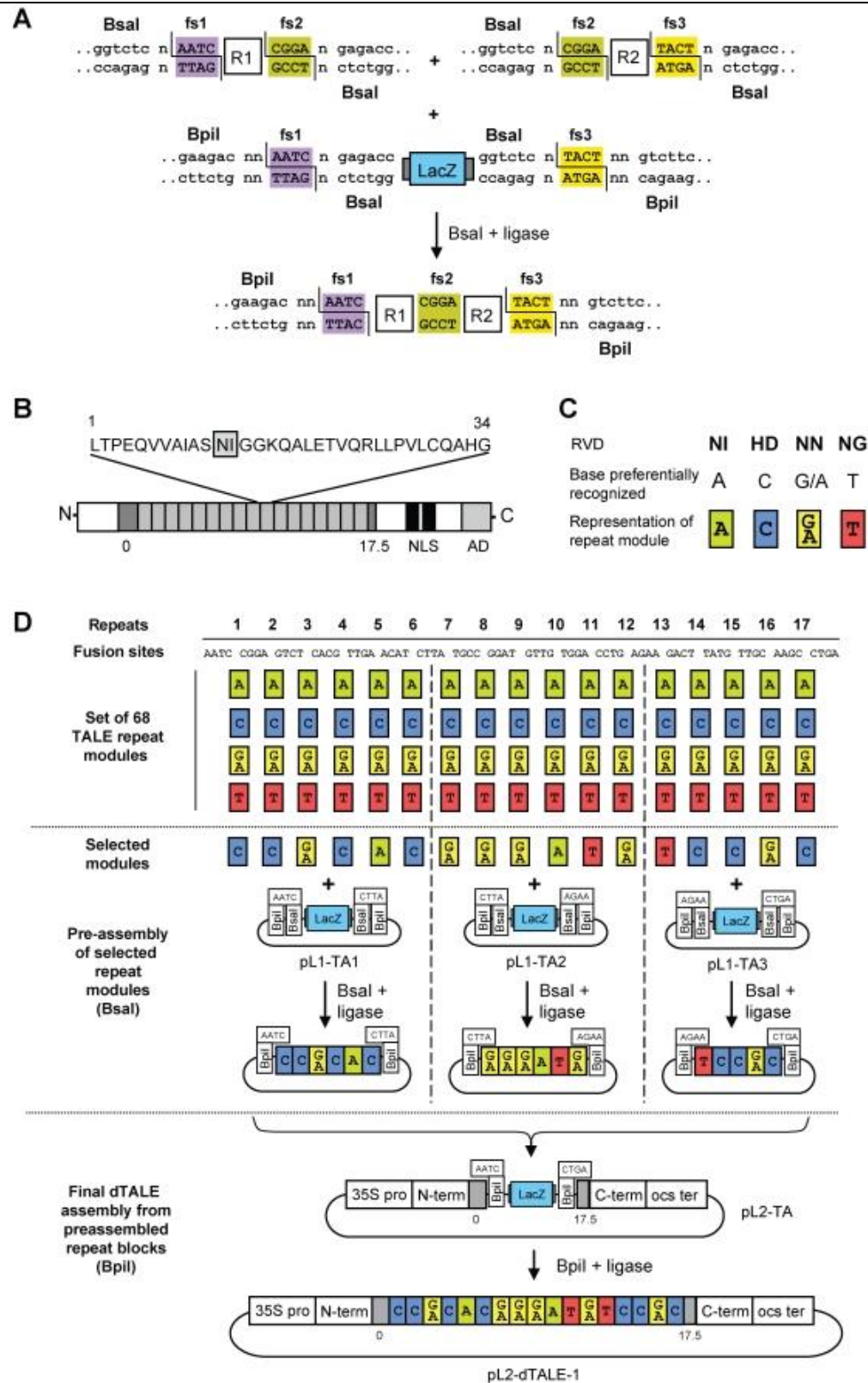


Figure 3 | General overview of the two-step cloning strategy for dTALEs assembly. A) Golden Gate cloning principle applied for assembly of dTALEs. Plasmids encoding selected repeat modules (an

example with only two modules, R1 and R2, is shown here due to space limitation) are mixed in one tube together with BsaI, T4 DNA ligase and the destination vector. Assembly of R1 and R2 using BsaI and ligase gives rise to a plasmid lacking the initial BsaI sites, but containing a block of assembled repeats flanked by two BpiI sites. The two BpiI sites allow release of the assembled repeats as one block for the second step of cloning. fs, fusion site. **(B)** Structure of AvrBs3. AvrBs3 contains a central region with 17 direct repeats (light grey boxes) flanked by a thymidine-specific repeat (repeat 0) and a half repeat (repeat 17.5, both flanking repeats shown as dark grey boxes). Two nuclear localization sequences (NLS, black bars) and a transcription activation domain (AD) are located in the C-terminal region. One representative 34 aa repeat is shown, with the RVD of the NI type highlighted in grey. **(C)** RVD types and their specificities. **(D)** Set of 68 repeat modules, with 4 modules with different specificities for each of the 17 repeat positions. Repeat modules are flanked by two BsaI sites with fusion sites selected from the codon-optimized sequence of AvrBs3. Sets of five (for repeats 13–17) or six (for repeats 1–6 and 7–12) selected repeat modules are preassembled via BsaI into preassembly vectors (pL1-TA1 to 3). Preassembled repeat blocks are then combined in the final destination vector (pL2-TA), using a second BpiI-based Golden Gate cloning reaction. Construction of dTALE-1 is shown as an example (10).

The length of spacer sequence is another limiting factor, especially in the TALEN system. The spacer sequence is a small fragment (12-25 bp) that separates the two TALEN monomers on opposite DNA strands, and cleavage by the FokI nuclease occurs here (1). If the spacer sequence does not have the proper size, it may prevent FokI dimerization (1, 4). Additionally, because of the large size and the highly repetitive nature of TALE effector DNA-binding domain, it is difficult to transfer TALEN by lentiviral vectors or adeno-associated viruses due the possibility of rearrangements in TALEN genes (1, 3). Overall, these are some of the challenges that are associated with using the TALEN system.

Besides these difficulties, targeting sequence for TALE has limitation, it should be a Thymine (T) before the 5'-end of the target site (1, 3). This could limit the cutting site of TALEN nuclease for breaking in any target region (1). Moreover, some studies revealed that the activity of TALE nucleases or recombinases greatly decreased in the absence of 5'T (11). However, some mutant variants of TALENs which are capable of binding with Adenine (A), Guanine (G), or Cytosine (C) can overcome this limitation (1).

Sometimes the engineered enzymes may cleave non-target sequences, result in off-target effects that are difficult to identify and are sometimes lethal (2, 5, 7). For instance, in ZFNs single nucleotide substitutions or inappropriate interaction between domains may occur (1). Therefore, some new systems and methods have been required to address these problems. CRISPR-Cas9

system was developed with new and unique features that has made it a more useful approach for gene editing (1, 2).

CRISPR-Cas9 was originally identified in bacteria and archaea which are used to defend against viruses and other foreign DNA (2). There are short repeats (CRISPR) in the bacterial genome which are separated by non-repetitive spacer sequences. These non-repetitive spacers are derived from foreign DNA originating from viruses. These CRISPR loci are transcribed into a long primary transcript which then is processed into a library of small CRISPR RNAs (crRNAs) (5) (Fig. 4a).

Each crRNAs contains a *guide* sequence complementary to a previously invading DNA and is bound with one or more CRISPR-associated (Cas) protein(s). When an invading nucleic acid or DNA enters into the bacteria, it is cut by Cas to short fragments and new spacers are added to the end of CRISPR locus. In the second exposure to foreign DNA, the bacteria will scan that genome for complementary sequences for the previous inserted sequences and then cuts invading DNA at that point, thus defending against foreign invaders (5) (Fig. 4a).

Scientists have utilized this newly discovered bacterial immune response as a gene editing approach in humans and cells of mammalian origin (2). They have synthesized short RNAs with complementary sequences of a target DNA. These RNAs are known as short guide RNAs or sgRNAs because they direct nucleases (usually Cas9) to the target DNA sequences (2, 6).

Each sgRNA is a single transcript that contains crRNA and trans-activating crRNA (tracrRNA) fused together (1-3). The crRNA has the *guide* sequence (20nt) and repeat sequence (12nt). The *guide* sequence pairs with target sequence on DNA strand and the repeat sequence attaches to one part of tracrRNA (anti-repeat). The tracrRNA also has a stem loop part that acts as a scaffold (2, 5, 6, 12, 13). After target sequence recognition and binding by sgRNAs, Cas9 will make a double-stranded cleavage at that specific site (2, 5, 6).

The double stranded break will repair by NHEJ or HDR repair pathways. If there is no repair template, cell starts to re-ligate DNA strand through NHEJ process. Some insertions or deletions may occur through this process which can result in gene knockouts, frameshift mutations or premature stop codons. While in the presence of an exogenous repair template, HDR pathway will start. The repair template can either be a conventional double-stranded DNA or single-stranded

DNA oligonucleotides. As a consequence, more precise and defined modifications will generate at a target locus than NHEJ repair pathway (14) (Fig. 4b).

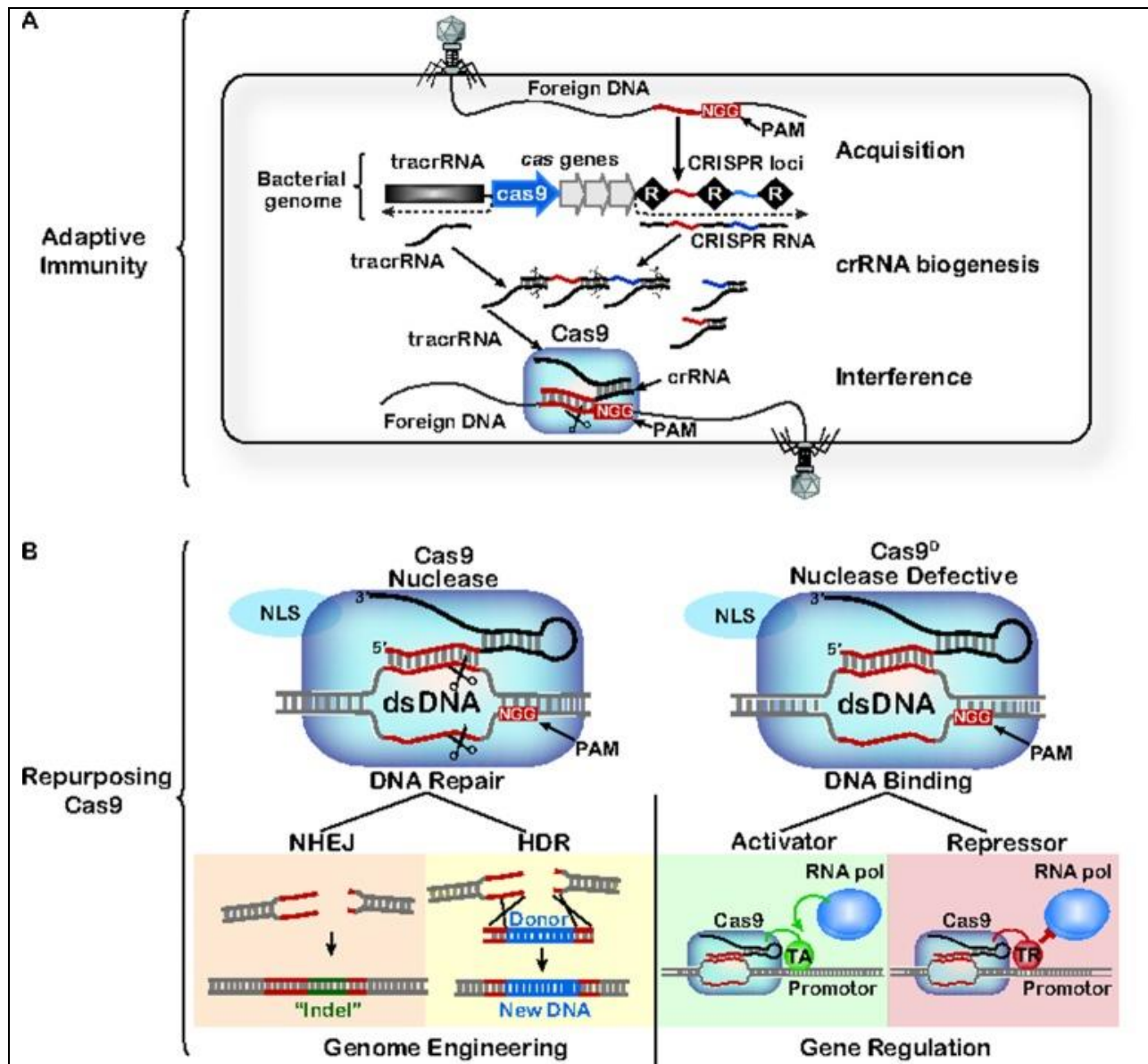


Figure 4 | Repurposing RNA-guided nucleases from the CRISPR-mediated adaptive immune system in bacteria. A) CRISPR-mediated adaptive immunity proceeds in three distinct stages: acquisition of foreign DNA, CRISPR RNA (crRNA) biogenesis, and target interference. Protospacers are short foreign DNA fragments which are integrated into the CRISPR loci encoded in the bacterial genome and make those bacteria resistance to the foreign nucleic acids. CRISPR loci consist of a series of short repeats (R, black diamonds) and unique spacers (red and blue lines). After CRISPR loci transcription, small CRISPR-derived RNAs (crRNAs) are formed. In some CRISPR systems (i.e. Type II), a trans-activating

CRISPR RNA (tracrRNA) is essential for RNA processing and Cas9 recognition. (B) Cas9 targeting relies on PAM recognition and base pairing between the crRNA and the target DNA. By designing a crRNA complementary to any target sequence with an adjacent PAM, Cas9 act specifically. Non-homologous end joining (NHEJ) or homology directed repair (HDR) are two repairing pathways which repair double-stranded breaks in DNA. Nuclease defective versions of Cas9 (Cas9^D) have been used to activate or repress transcription of target genes (5).

All of the methods mentioned above have been useful for gene editing (1-3). In the next chapter, the work of Parnas and his colleagues will be reviewed to show the efficiency and advantages of CRISPR-Cas 9 system for finding new regulatory genes in the tumor necrosis factor- alpha (TNF- α) induction pathway (15). This review also will describe the features of this system which made it more advanced than previous tools with regard to its accomplishments. The challenges will be acknowledged, and suggestions to address some of the challenges will be proposed for further researches.

Chapter 2: Methodology

The main article *A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks* was given by my supervisor Doctor Ron Shigeta. Utilizing the Pubmed database, a combination of search keywords including "CRISPR-Cas9 procedure", "Cas9", "TNF induction pathway", "Bacterial LPS", "CRISPR-Cas9" and "Advantages", "TALEN", "ZFN", and "CRISPR-Cas9 accomplishments" were used. Research articles were prioritized and some review papers were also selected.

Chapter 3: Evaluating Parnas et al. Research

Although finding the components of the cellular circuit and determining their functions in mammalian cells remains a challenging issue, Oren Parnas and his colleagues investigated regulatory genes in TNF induction pathway after bacterial lipopolysaccharide (LPS) exposure by new advanced CRISPR-Cas9 system (15).

3.1. TNF induction pathway and its components

TNF- α is a mediator that is produced by immune cells and can play a key role in inflammatory responses (16, 17). As a byproduct of the inflammatory signaling cascade, TNF- α can activate T cells, regulate macrophage functions and the production of other downstream inflammatory cytokines (18, 19). TNF- α is also one of many cytokines that can be critical for sustained inflammation. For example, an in vitro study on macrophages investigated an autocrine loop that was induced by TNF- α during a 2-day period after TNF- α stimulation. In this loop, TNF- α and Interferon beta (IFN- β) act in parallel result in delayed expression of interferon-response molecules' genes and prevent the toxicity of high interferon production (19).

TNF- α also has functions in immunity and immune regulation. For example, it participates in the development of follicular dendritic cells, secondary lymphoid tissue and generation of primary B-cell follicles. It can inhibit apoptosis or programmable dead of cells by inducing a variety of anti-apoptotic molecules such as TNF-receptor associated factor. These effects of TNF- α besides its inflammatory role are also evident in autoimmune diseases such as human systematic lupus erythematosus (SLE) (20).

SLE is an autoimmune disease with antibody production by our immune system against ourselves. TNF- α level increases in SLE. As TNF- α is a growth factor for B cells, significant amounts of B cells and autoantibodies are produced. Moreover, several data suggested that tissue destruction and inflammation in SLE are because of immune complexes (antigen-antibody) formation, macrophages activity, and production large amount of TNF- α in those tissues (20).

When our body is infected with gram negative bacteria that express LPS on their cell surface, TNF- α is the first major host-mediated defense response (16, 21). LPS begin its function by its receptor. If one cell has receptor for LPS, it will bind with its receptor, activate immunity and form the LPS-binding protein (LBP) (16).

The first host protein involved in LPS recognition is LBP. It recognizes and forms a high-affinity complex with the lipid “A” residue of LPS. This helps LPS form a ternary complex with a co-receptor known as Cluster of Differentiation 14 (CD14). CD14 recognizes LPS-LBP complex and presents it to another receptor, lymphocyte antigen 96 or MD-2. The secreted glycoprotein, MD-2, acts as an extracellular adaptor protein in the activation of toll-like receptor 4 (TLR4) by LPS and is essential for LPS signaling to occur (22).

LPS binds with MD-2, which in turn associates with another receptor, TLR4. Thus, a series of signaling pathway known as TLR signaling pathway occurs. Several proteins and transcription factors are activated through this pathway such as myeloid differentiation factor 88 (MYD88) (22). As a result, large quantities of cytokines and immune-relevant proteins are produced. TNF- α , interleukin-1 (IL-1), and IL-6, or IL-10 are some of the cytokines that can be produced in response to LPS (23) (Fig. 5). Subsequently, TNF- α is released and some genes regulate its production positively and/or negatively (15).

TNF- α production, its amplitude and duration of its function needs to be tightly regulated (24). One example of the consequences of TNF- α dysregulation is septic shock. This occurs in response to LPS signaling during infection with gram-negative bacteria. If TNF- α production does not regulate, large amounts of inflammatory cytokines and molecules will be produced. This results in septic shock's symptoms such as fever, hypotension, and multiple organ failure (25). Autoimmune diseases such as rheumatoid arthritis (RA) is another example of TNF- α dysregulation outcome (26). Thus TNF- α production and inflammation should be regulated (24).

Because of the TNF- α dysregulation outcome, TNF pathway has this ability to regulate itself. For example, some findings revealed that TNF- α has a major effect on IL-10 mRNA production and increases its production. IL-10 will down-regulate the inflammatory cytokines production such as

TNF- α as well as itself. Furthermore, TNF- α stimulates a population of T cells which then inhibit further TNF- α synthesis (18, 27, 28).

These CD4⁺FoxP3⁺ T cells are called regulatory T cells (Tregs). Tregs target a wide range of immune cells such as CD4 cells, CD8 cells, natural killer (NK) cells, dendritic cells, macrophages and B cells to suppress them and maintain immune responses with the contribution of several molecules such as IL-10. Studies show the regulatory and pro-inflammatory effects of TNF- α by a negative feedback loop (18, 27, 28).

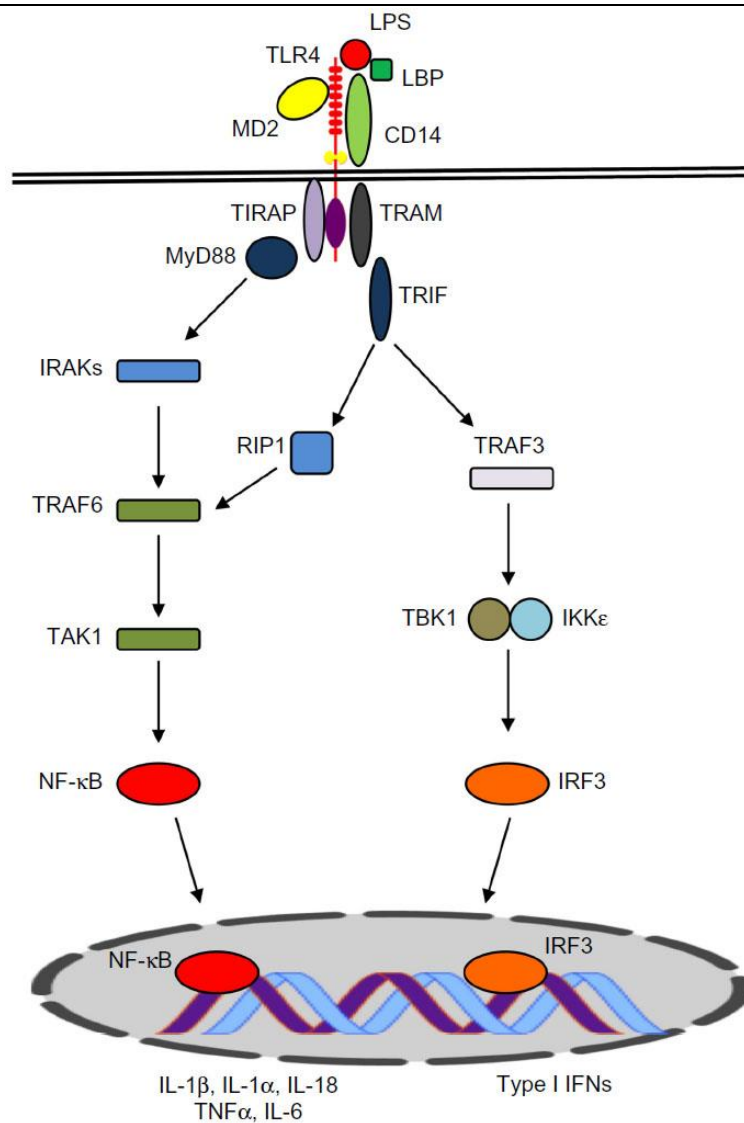


Figure 5 | Schematic view of the TLR4 signaling pathway. Upon binding of lipopolysaccharide (LPS) to the TLR4 complex (TLR4–LBP–CD14–MD2), TLR4 signals through MyD88-dependent and/or

MyD88-independent pathways. The MyD88-dependent pathway is triggered by MyD88 recruitment of TRAF6 and IRAKs, which, in turn, results in activation of TAK1. TAK1 phosphorylates IKK β , leading to I κ B α and activation of NF- κ B. NF- κ B translocates to the nucleus to promote the transcription of proinflammatory genes (such as IL-1 α/β , IL-18, IL-6, and TNF α). The MyD88-independent pathway relies on TRIF recruitment of RIP1 or TRAF3. TRAF3 activates IRF3 through TBK1, inducing transcription of type I interferons (IFNs) and IFN-inducible genes (29).

3.2. Cells and mice preparation

In the Parnas study, bone marrow dendritic cells (BMDCs) were used as primary immune cells to investigate TNF- α signaling pathways (15). Perhaps they assumed that dendritic cells are antigen-presenting cells that are involved in the generation of host-responses against pathogens or altered self (30). In addition, these cells are one of the “professional” and powerful antigen-presenting cells in the inflammatory response (16).

First, researchers constructed transgenic mice that express Cas9 (15). Transgenic mice are kinds of laboratory animals which are important for modeling human diseases and testing therapeutic strategies for removing or modifying genes in mouse genome (Knockout or knock-in mouse) (31). As Parnas and colleagues aimed to find regulatory genes in mammalian cells, mice could be a good choice. Cas9 expressing transgenic mice genetically modified to express this endonuclease gene stably (32).

The first step for generating transgenic mice is to obtain genes of interest and grow many copies of them in vectors. In the next step, only desired genetic fragments have to be separated completely from vector backbone by proper restriction enzymes. After that, many copies of the transgenes need to be microinjected into the semi-fertilized egg and then these eggs should be implanted into a pseudo pregnant female mice. Finally, the first progeny has to be genotyped and analyzed for transgene expression (33).

Mouse lines with Cas9 have not shown observable phenotype after Cas9 incorporation into their genomes (34). Although, it has been shown that prolonged expression of Cas9 and sgRNA can

significantly increase the knockout efficiency, it has been amplified off-target effects. Off-target effects may be reduced by transient expression of Cas9 (35).

3.3. Making sgRNA Pooled Libraries

In the second step, a pooled library of sgRNAs was made (15). sgRNAs can be synthesized quickly with a very low cost for multiple targets (36). Making sgRNA library for CRISPR-Cas9 system requires different techniques.³²The molecular chipper method generates large arrayed sgRNA libraries in four rounds of assembly and sequencing (37, 38).

In the first step input DNA pieces were fragmented with restriction enzymes, resulting in randomly distributed fragments. As the targeted sequences must be approximately 20 nucleotides in length and immediately 5' to a protospacer adjacent motif (PAM) sequence, the restriction enzyme should target the 5' end of DNA to NGG or NAG sequences. The resulting fragments then were ligated to an adaptor containing a recognition site for the restriction enzyme which cuts 20-21 nt 5' (figure 6) (37, 38).

In the next step, these fragments were ligated with the non-targeting portion of sgRNA and finally cloned into viral sgRNA expression vector (Fig. 6b). For sgRNA expression, the U6 promoter was put in the vector. The U6 promoter is important for eliminating post-transcriptional modifications on sgRNAs such as 5' methyl capping and polyadenylation (13). They also used cytomegalovirus (CMV)-driven Green Fluorescent Protein (GFP) gene in the vector backbone for transfection or transduction control. Then by a two-step polymerase chain reaction (PCR), they attached Illumina-compatible linker sequences and three barcodes to identify all members of each pool by deep sequencing (37, 38).

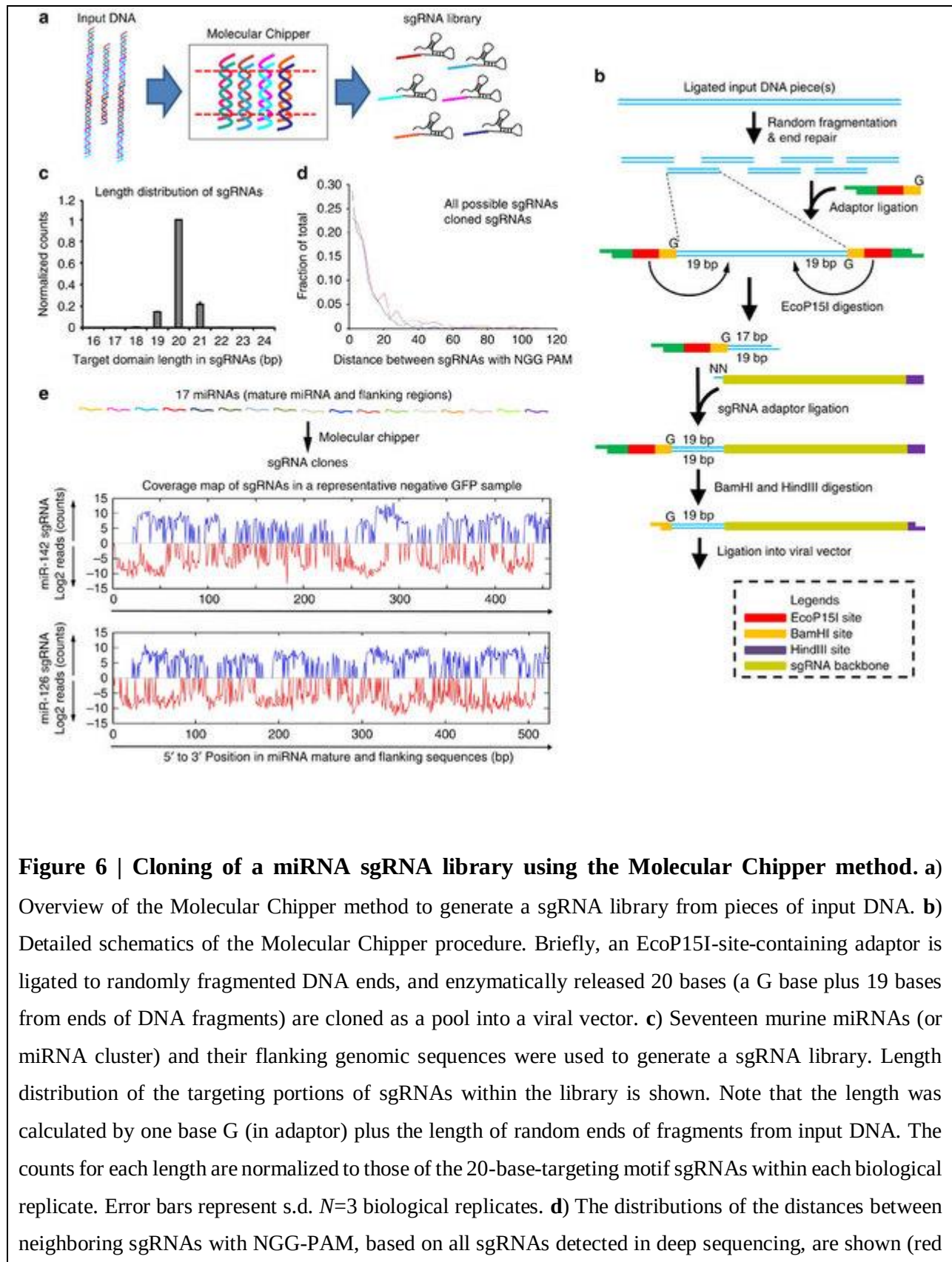


Figure 6 | Cloning of a miRNA sgRNA library using the Molecular Chipper method. a)

Overview of the Molecular Chipper method to generate a sgRNA library from pieces of input DNA. **b)**

Detailed schematics of the Molecular Chipper procedure. Briefly, an EcoP15I-site-containing adaptor is ligated to randomly fragmented DNA ends, and enzymatically released 20 bases (a G base plus 19 bases from ends of DNA fragments) are cloned as a pool into a viral vector. **c)** Seventeen murine miRNAs (or miRNA cluster) and their flanking genomic sequences were used to generate a sgRNA library. Length distribution of the targeting portions of sgRNAs within the library is shown. Note that the length was calculated by one base G (in adaptor) plus the length of random ends of fragments from input DNA. The counts for each length are normalized to those of the 20-base-targeting motif sgRNAs within each biological replicate. Error bars represent s.d. $N=3$ biological replicates. **d)** The distributions of the distances between neighboring sgRNAs with NGG-PAM, based on all sgRNAs detected in deep sequencing, are shown (red

line). The median neighbor distance is 8 bp. Theoretical distribution assumes all possible NGG-PAM sgRNAs (blue line) are present. **e)** Top: diagram showing that the 17 murine miRNAs (or miRNA cluster) and their flanking genomic sequences were used to generate a sgRNA library. Bottom: representative graphs of sgRNA counts mapping to the miR-142 region or to the miR-126 region from one out of three neg-GFP samples is shown, with blue and red indicating mapping to sense and antisense strands, respectively. The positions of sgRNAs plotted were only based on positions of the last targeting domain base (38).

3.4. Delivery of CRISPR-Cas9 Components

The other necessary tool was lentiviral vectors because of several reasons (15). First, this kind of vector are employed for large-scale genome-wide knockout and using large sgRNA libraries (1). As Parnas and his colleagues had to deliver 125,793 sgRNAs into the target cells, they needed this kind of vector (15). The other unique advantage of Lentiviral (LV) vectors over other gene delivery systems is that they have the ability to integrate transgenes into the genome of both dividing and non-dividing cells as they are stable genomic integrant during cell replication (39).

Lentiviral vectors have this ability to deliver Cas9 and sgRNA simultaneously. It will enable CRISPR-Cas9 to be used in any cell type of interest and eliminate the need to generate cell lines that express Cas9 (7). In this study, Parnas et al generate Cas9 mouse cell line beside lentiviral vectors (15). If they deliver Cas9 and sgRNA through a single vector, it will not need to generate Cas9 cell line, and therefore, it may increase their speed of experience and reduce the cost of research.

All sgRNA expression plasmids were assembled by designing, synthesizing, annealing, and cloning pairs of oligonucleotides (IDT) harboring the complementary region into vector (40). The first step for lentiviral vectors assembly is the annealing of single stranded oligos containing each 20bp protospacer to create overhangs for ligation into the desired donor promoter vector. Protospacer is ligated into the desired expression vector by using T4 DNA ligase (41).

In the second step, independent plasmids, which each expresses a different sgRNA from a different promoter, transfer into lentiviral destination vector using Golden Gate assembly and using T4 DNA ligase. In the last step, desired cells are transfected with these viral vectors and then plated

to form colonies. Single cell clones which express GFP (reporter gene co-expressed with Cas9 to facilitate quantification of transduction and/or fluorescence-based sorting of transduced cells) are isolated and genomic DNA are harvested for analysis. For viral transduction, cells are plated in a dish and the viral supernatant is added (41).

3.5. sgRNA Abundance Determination

Deep sequencing is another method which was used for determining sgRNA abundance (5). Deep sequencing or Next Generation Sequencing (NGS) is a technique to detect genome modifications but it is suitable for sampling a large number of samples or target sites. For this method, primers are designed that are complementary to part of DNA sequence. For the detection of NHEJ mutations, it is important to design primers situated at least 50 bp from the Cas9 target site to allow for the detection of longer mutations. For larger deletions mediated by multiple sgRNAs, priming sites should be designed outside the deleted region (14).

Deep sequencing costs can be reduced by adding *barcodes* to individual libraries so that they can be mixed together and sequence in a single lane. *Barcodes* consist of 2–10 unique nts appended either 5' or 3' to the captured sequences, and ideally differ by more than 2 nts so as to minimize incorrect library identification due to sequencing errors (42).

The NGS technique has several steps that include: 1- preadenylated 3' adaptor ligation, 2- reverse transcription, 3- gel purification, 4- circularization, and 5- PCR. In the first step, single-stranded RNA and 3'-adaptor are mixed and then the RNA ligase are added. In the second step, RNA from the first step incubated with reverse transcriptase (RT) primer and reverse transcriptase enzyme to transcript RNA into circle DNA (cDNA). To maximize RT product yield, it is important that the RT primer concentration be greater than the 3'-adaptor concentration but not excessively, as this would favor empty circle formation in the subsequent circularization reaction (Step 4) (42).

Reverse transcription samples are diluted, denatured, and then analyzed on denaturing polyacrylamide gel electrophoresis (PAGE) gels in the third step (39). In step 4, all single-stranded DNA (ssDNA) isolated in step 3 are incubated with CircLigase enzyme, and in the end, circularized RT product are amplified by the PCR method. After that all PCR products are analyzed on PAGE gels (42).

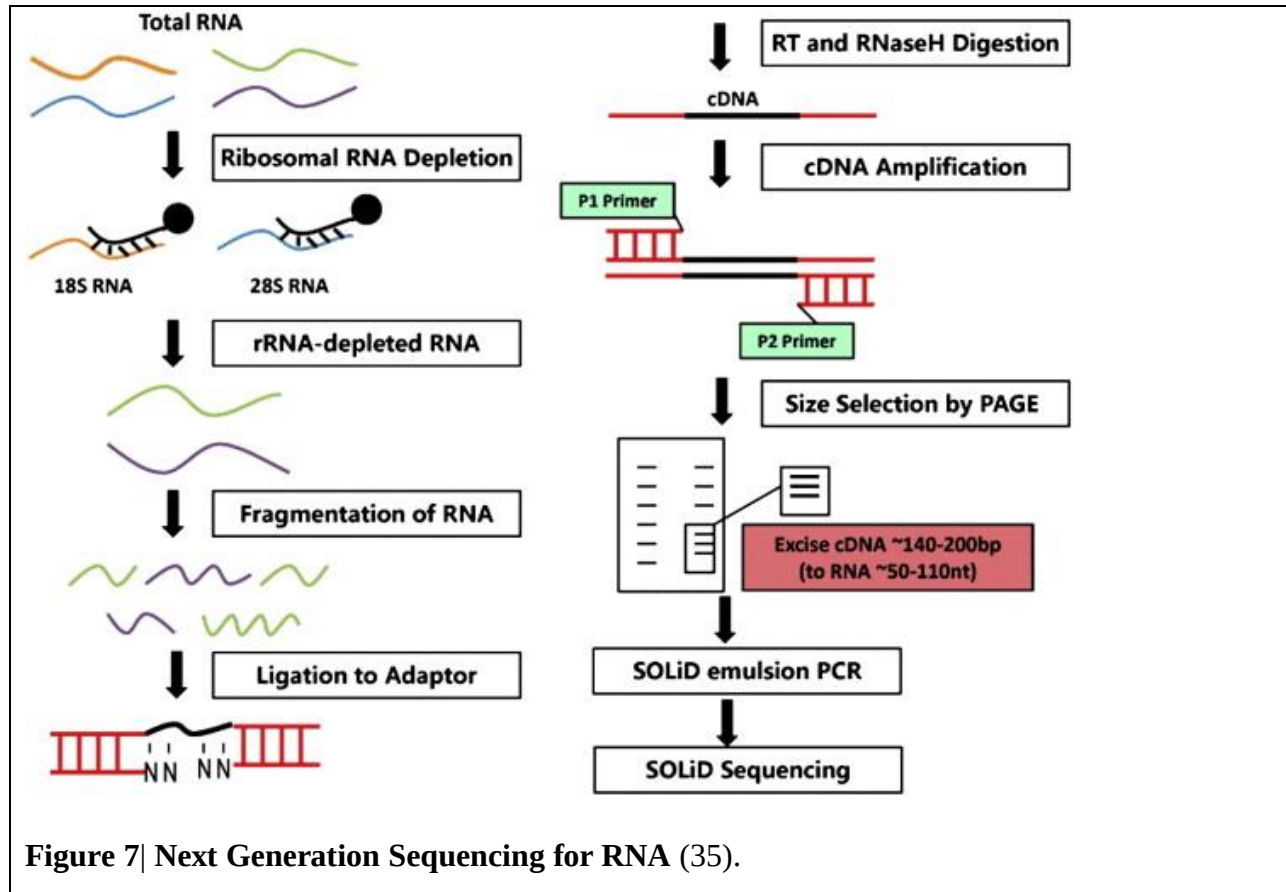


Figure 7| Next Generation Sequencing for RNA (35).

3.6. Results

As a result, many of the key known regulators of TLR signaling have been recovered, dozens of new regulators have been validated, and three functional modules of regulators with distinct regulatory effects have been identified. These modules were based on regulators' impact on RNA and protein expression of selected markers of dendritic cell function (15).

In this experiment, lentiviruses expressing sgRNA targeted each of three genes:

- 1- TLR4, the cell membrane receptor that senses bacterial LPS
- 2- MYD 88, a key component required for TLR4 signaling to induce TNF- α . MYD 88 level increased after LPS exposure (21-43)
- 3- ZFP36 (TTP), zinc finger protein that is a member of the tristetraprolin (TTP) family, an RNA-binding protein that destabilizes TNF- α mRNA (44)

These results were compared with non-targeting sgRNA as a control (15).

The first screen performed by infection of 60-200 million BMDCs with a library of lentiviruses harboring 125,793 sgRNAs targeting 21,786 annotated protein-coding and miRNA mouse genes, as well as 1000 non-targeting sgRNAs as negative controls. These cells which stimulated with LPS then were sorted based on high or low TNF- α expression levels (15). sgRNAs were extracted by specific reagent, amplified by specific primers and PCR technique (21) and then sequenced at the next stage from four sources:

- 1- Post-LPS cells with high TNF- α
- 2- Post-LPS cells with low TNF- α
- 3- Cells from the last day of differentiation prior to LPS stimulation
- 4- Plasmid DNA of the input lentiviral library (15).

The main purpose for this screening was to discover both positive and negative regulators of TNF- α induction after LPS exposure, but they concluded that CRISP-Cas9 screening was less sensitive for negative regulators at high (near-saturation) levels of LPS.¹¹ To increase sensitivity for negative regulators, they reduced LPS by 5-fold that led to the discovery of new sets of TNF- α regulators (15).

Another innovation was secondary dipper screen for improving specificity and sensitivity, comprising the top ranked 2569 genes. For this, they increased the number and efficacy of sgRNAs per genes and the number of cells infected per sgRNA. On the basis of these screening, targeting genes were partitioned into three major modules (15).

Module I consisted of sgRNAs targeting 17 genes including 9 canonical regulators validated in the screen, each reducing the levels of Cd14 and Il6, but not Cd11c, consistent with the roles of the known regulators in LPS signaling (15) After LPS exposure, CD14 makes complex with LPS which couple with TLR-4-MD-2 to activate cells. MyD88 also participates in cell activation by TLR-4 (45).

Module II included nine regulators whose sgRNAs reduced all four proteins (15). Among them, were four subunits of the oligosaccharyltransferase (OST) protein glycosylation complex such as Alg2. It is a glycosyltransferase involved in oligosaccharide synthesis by catalyzing the glycan

transfer to polypeptide chains (46), and genes whose molecular functions are currently unknown, such as Tmem258 (15).

Module III consisted of sgRNAs targeting three subunits of the RNA polymerase II-associated factor (PAF) complex and Pol2rg (15). Each reduced Cd11c and Il6 expression (47, 48), because PAF complex can interact with histone-modifying enzymes and RNA polymerase II to regulate transcriptional elongation (48). The PAF complex was found in the promoter regions of IL-6 and so reduced IL-6 transcription but had a very minor, albeit consistent, effect on Macrophage inflammatory protein-1a (Mip1a) that can be secreted by macrophages and no effect on CD14 (49).

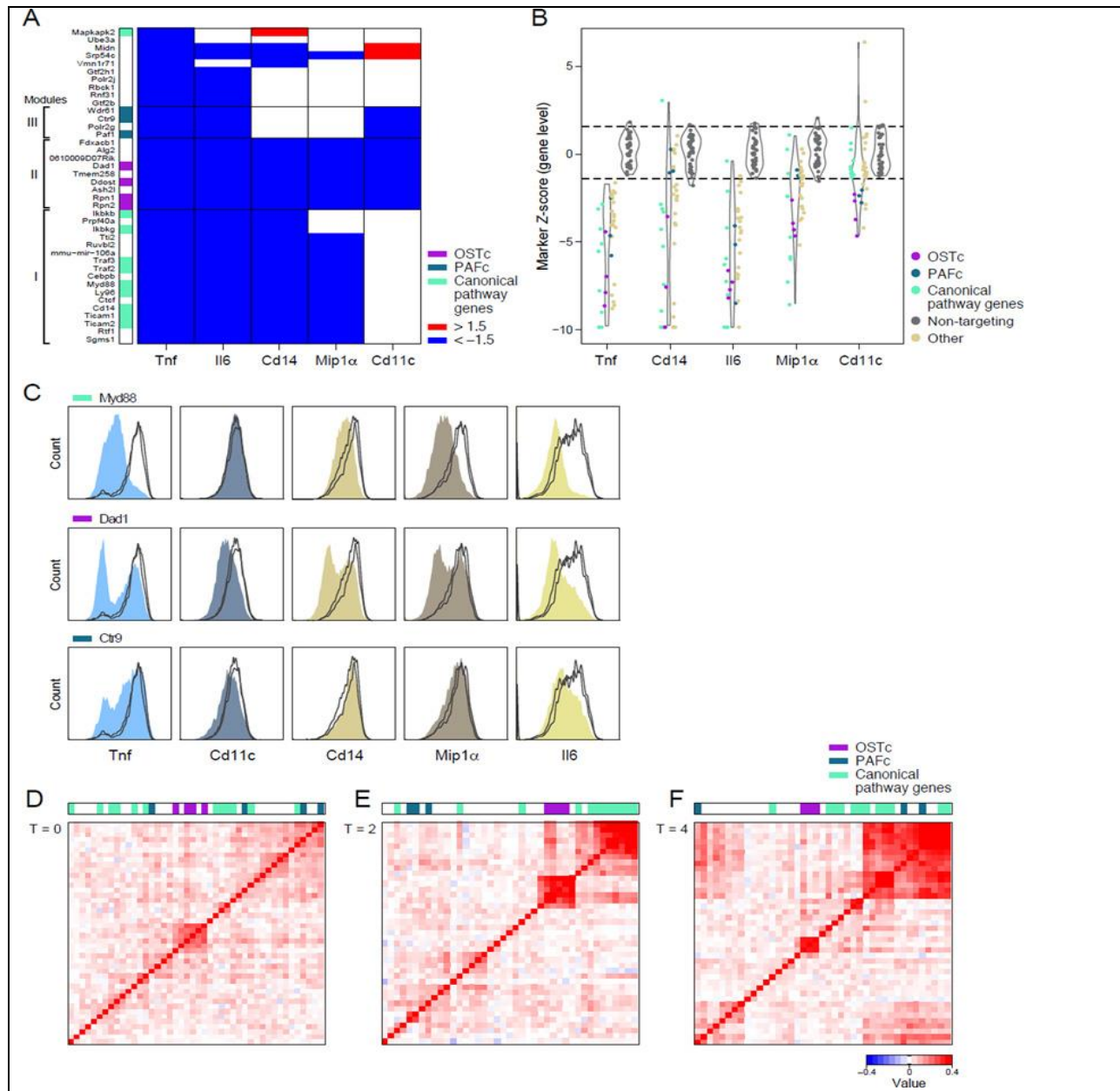


Figure 8 | The validated positive regulators partition into key modules by their effect on protein and RNA expression. A) Change in expression (blue, reduced; red, increased) of 5 protein markers (labeled columns) measured by flow staining with antibodies for cells with sgRNAs targeting the indicated genes (rows). Three modules indicated with brackets, and color bar on left corresponds to legend on right. **B)** Violin plots of the distribution of Z scores of true positive regulators of Tnf (left) or of non-targeting control sgRNAs (right) for each marker. Functional groups are colored as in (A). **C)** Effects of selected sgRNAs targeting genes in each of three modules on protein markers for true positives (filled) vs.

non-targeting controls (lines). **D-F**) Correlation of global RNA expression profiles (normalized to non-targeting control values) for verified positive regulators per time point post-LPS, as indicated (15).

3.7. Some Points Regarding the Authors' Research

As studying regulatory networks and finding the components of cellular circuits and determining their functions are major challenges in mammalian cells (15), what Parnas et al have done is worthy and valuable. Their study showed the potential of CRISPR-Cas9 in exploring functional networks and suggests new ways for further research.

The authors tried to improve accuracy, sensitivity and specificity in all stages. For instance, they added Brefeldin A to block TNF- α secretion (15). Brefeldin A is an inhibitor of intracellular protein transport from the endoplasmic reticulum (ER) to the Golgi apparatus. Adding Brefeldin A could have another positive effect on CRISPR potential as it could improve the HDR efficiency and insertion efficiency (50). Thus could decrease mutations and improve sensitivity.

Parnas et al performed a secondary screen with higher sensitivity and specificity due to increased amount of sgRNAs per genes and cell numbers per sgRNAs. If the number of cells increase, it might increase off-target effects and decrease sensitivity of the experience. In addition, they determined sgRNA abundance two times for each top ranked genes to gain more sensitivity. At the end of this stage, new regulatory genes were identified (15).

The other important point about this research is the detection of false-positive and false-negative due high numbers of cells regarding the size of sgRNA library. The authors addressed this issue by decreasing the target genes and using 10 sgRNA per gene and 4.9 fold cells per sgRNA. Subsequently, more known regulators and true positive genes and 19 new positive regulators were identified (15).

Although the existence of false-positive and false-negative may create doubts regarding CRISPR potential, this research showed that CRISPR-Cas9 system can aid in the discovery of novel regulatory pathways. As a positive control, all known regulators of the TLR dependent TNF- α signaling pathway were identified correctly (12) and the results were consistent with the authors' predictions (15). Whether any of the newly identified negative regulators are involved in any diseases remains to be determined.

Chapter 4: Advantages of CRISPR-Cas9 System

The CRISPR-Cas9 system has some potentials more than other genetic engineering systems that make it a powerful and multiplexable genome editing tool. Comparing with ZFNs and TALENs, which are built on protein-guided DNA cleavage and need complex and time-consuming protein engineering, selection and validation, CRISPR/Cas9 only needs a short programmable sgRNA for DNA targeting, which is relatively cheap and easy to design and produce (51).

By using Cas9 and several sgRNAs with different target sites, CRISPR-Cas9 is able to simultaneously induce genomic modifications at multiple independent sites. This technology can accelerate the generation of transgenic animals with multiple gene mutations, and disrupt multiple genes or a whole gene family to investigate gene function and relationships (51).

4.1. Cas9 Properties and Potentials

Cas9 can make double stranded breaks in DNA (14, 51-53). This ability of Cas-9 depends on the guide RNA's function. crRNA arrays are short sequences derived from exogenous DNA known as protospacers. Each crRNA unit contains a 20-nt guide sequence and a partial direct repeat. This *guide* sequence directs Cas9 to a 20-bp on the target DNA via base pairing that is complementary to protospacer (14, 51). By altering the 20-nt guide sequence within the sgRNA, Cas9 will be directed towards almost any target of interest (14).

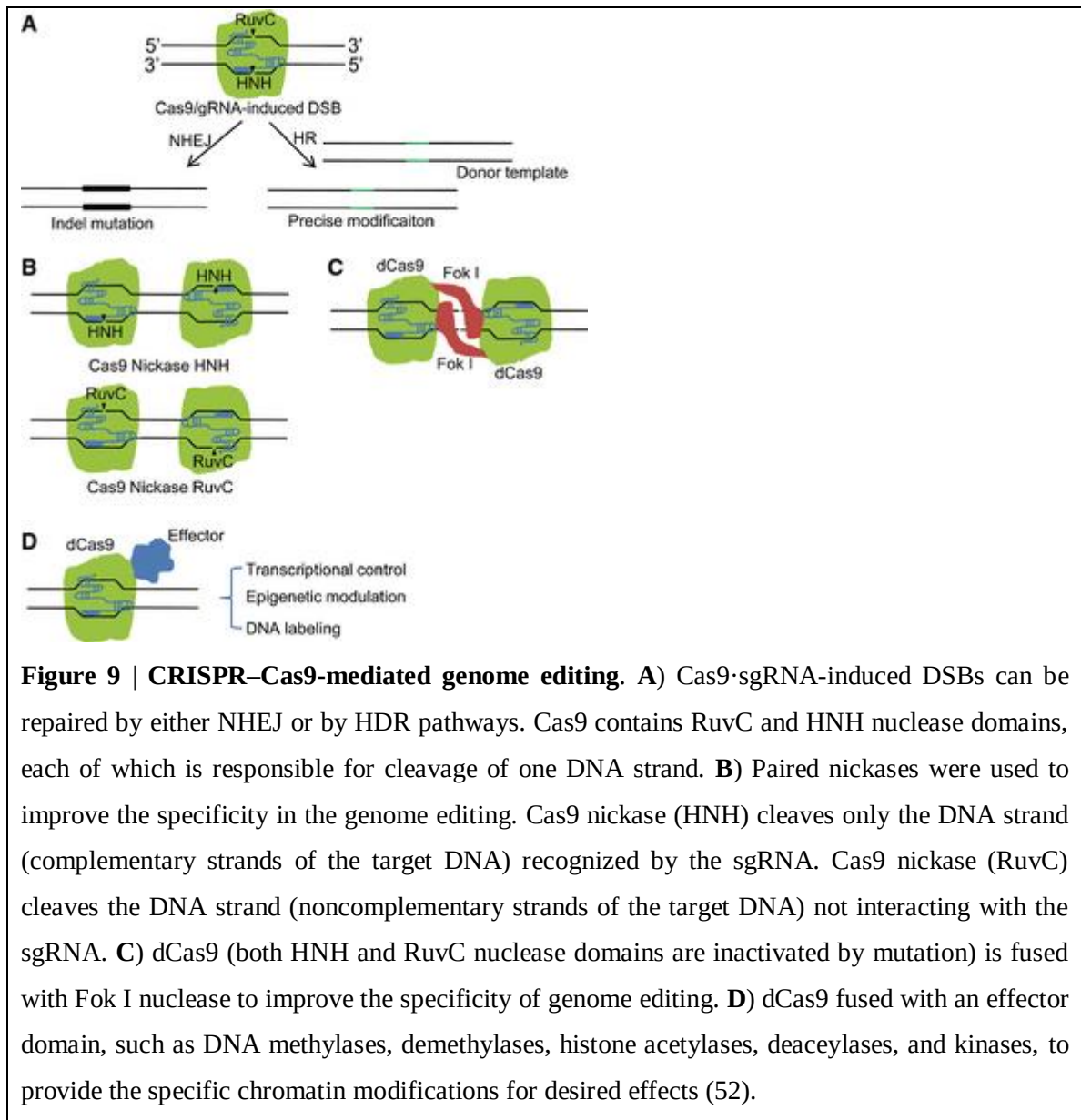
The other necessary element for Cas9 recognition is PAM which are NGG motifs and are located immediately downstream of the target sequence. The identified PAM sequences vary among different Cas9 orthologs. Indeed, the specificity of CRISPR/Cas9 is determined partly by this 2–5 nt PAM sequence (51-53).

Knowing about the conformation of Cas9 and its interaction with RNA and DNA during RNA-guided DNA cleavage have made the mechanism of Cas9 more understandable. At first, Cas9 binds to the target DNA by PAM recognition and after a conformational change and RNA addition to the Cas9, an R-loop structure is formed. In R-loop conformation, crRNA can bind to target region of complementary DNA strand and then no noncomplementary strand (54).

Additionally, the length of base pairing between crRNA and target DNA affects the cleavage activity of Cas9. 8 nt base pairing is necessary for sgRNA-Cas9 binding to the target strand. Although 12 bp is the minimal size for the active conformation of Cas9 (R-loop) and nicking of the target strand, simultaneous cleavage of both DNA strand occurs when the length of base pairing between sgRNA and DNA is 14-17 bp (54).

Cas9 also has another property that can facilitate eukaryotic gene editing. This endonuclease contains the RuvC and HNH nuclease domains. The HNH domain is a single domain, whereas the RuvC domain consists of three subdomains. These two domains are responsible for the cleavage of the complementary and noncomplementary DNA strands of the target sites (52, 53).

The RuvC and HNH domains can be changed and exploited for additional functions (40). For instance, a change in one amino acid in the RuvC catalytic domain generates a new variant of Cas9 as Cas9 nickase (Cas9n). Cas9n is able to nick rather than cleave DNA to yield single-stranded breaks. The subsequent repair through HDR can potentially decrease the frequency of unwanted mutations. Additionally, sgRNA pairs can guide Cas9n to simultaneously nick both strands of the target locus to mediate a double nicking. This can reduce the likelihood of off-target modifications and thus increase the specificity of target recognition (14).



4.2. Other advantages of Cas9

- **Ease of customization.** Cas9 can be easily retargeted to new DNA sequences by simply purchasing a pair of oligos encoding the 20-nt guide sequence. In contrast, retargeting of TALEN for a new DNA sequence requires the construction of two new TALEN genes. Although a variety of protocols exist for TALEN construction, it takes substantially more hands-on time to construct a new pair of TALENs (14).

- **Cleavage pattern.** Wild type *Streptococcus pyogenes* Cas9 (SpCas9) is known to make a blunt cut between the 17th and 18th bases in the target sequence (3 bp 5' of the PAM). Mutating catalytic residues in either the RuvC or the HNH nuclease domain of SpCas9 converts the enzyme into a DNA nicking enzyme. In contrast, TALENs cleave nonspecifically in the 12–24-bp linker between the pair of TALEN monomer-binding sites (14).
- **Editing efficiency.** SpCas9 and TALENs have both been shown to facilitate efficient genome editing in a variety of cell types and organisms. However, owing to the ease of targeting, Cas9 can be used to target multiple genomic loci simultaneously, by co-delivering a combination of sgRNAs to the cells of interest (14, 41).

Recently, a new system has been developed that permits multiple expression of Cas9 or a dCas9 variant with two or more sgRNAs and a reporter gene, from a single lentiviral vector (41, 55). In this plasmid each gRNA flanked by cleavage sites for a ribonuclease (for releasing gRNAs in a cell), is expressed within a single RNA, and transcribed by a single Human U6 promoter (55). Each promoter is capable of efficiently expressing sgRNAs that directs similar levels of Cas9 nuclease activity. Furthermore, lentiviral delivery of a single vector expressing Cas9 and multiple sgRNAs targeting distinct loci results in simultaneous multiple gene editing of all targeted loci (41, 56). This ability not only enhances the editing efficiency but also ensures gene knockout through genomic deletion (56).

Cas9 has this potential for diverse applications, such as recruitment of fluorescent protein labels or chromatin- modifying enzymes to specific genomic loci for reporting or modulating gene function (14). It is also possible to fuse effectors directly to its mutated version and convert it into a programmable artificial transcription factor (53). This could add other abilities to Cas9. For instance, dCas9-VP64 is able to activate silent endogenous genes and reporters or upregulate already active genes (53).

4.3. Activating and Repressing Genes by CRISPR-dCas9

Regulation of each gene or loci in the genome is one of the important and sometimes vital processes in organisms. For example, if one cell wants to have new fate during development, it is critical to activate some genes which are associated with its new fate and repress sets of other

genes that are related to its prior or alternative fate. Sometimes, cells are going under some environmental pressures that lead to some changes in their metabolisms. Activating of some new genes or enzymes and repressing other previously expressed enzymes is necessary for tolerating new conditions. Regulators such as transcriptional activators or repressors are required for gene regulation (56).

Transcriptional activators are some factors which bind to specific sequences upstream of genes, such as promoter or enhancer, then facilitate binding and moving of RNA polymerase. Transcriptional repressors are those factors which prevent the RNA polymerase (RNAP) from binding to promoter and thus repress transcription (57). Some of the activators are like herpes simplex virus VP64 activation domain (58), omega subunit of RNAP (57), nuclear factor NF-kappa-B p65 subunit (p65) or (RelA), the Epstein–Barr virus R transactivator (Rta), the catalytic histone acetyltransferase core domain of the human E1A-associated protein p300 (dCas9-p300) (59), and repressors are such as the Kruppel associated box (KRAB) repressor domain of KOX1, the chromoshadow (CS) domain of HP1a, the WPRW domain of Hes1 (60) or the EAR-repressor domain (SRDX) from Arabidopsis SUPERMAN protein (58).

CRISPR-Cas9 can also participate in gene regulation by requiring transcriptional factors to regulatory sequences upstream target gene or genes. For switching to the application of CRISPR-Cas9 to a regulatory system, nuclease activity of Cas9 has to deactivate (57, 60). By a substitution of aspartic acid (D) to alanine (A) at the 10th amino acid in the RuvC-like domain and histidine (H) to alanine (A) at the 840th amino acid in the HNH domain of Cas9, Cas9 will lose its nuclease function while preserve its binding affinity to DNA and is named as dCas9 (D10A/ H840A) (58). dCas9 has also this ability to fuse with activators or repressors and acts as transcription factor (58, 60).

CRISPR interference or CRISPRi is the modified system that block transcription and repress gene expression (61). When dCas binds to promoter, it prevents RNAP binding to promoter and represses transcription. dCas9 can also bind to open reading frames and block transcription elongation (57). Additionally, binding of transcription repressors to dCas9 improves the activity of dCas9 in disrupting RNAP activity and so efficiency of repression especially in mammalian cells (60).

CRISPR activator or CRISPRa is another type of CRISPR method in which transcription activators are fused to dCas9 for enhancing transcription of target genes. The basis of activation is on directing RNAP to promoter and facilitating its binding (57).

Several studies have been done to show the efficiency of CRISPRi on repressing prokaryotes and eukaryotes genes. For instance, binding of dCas9 with the help of crRNAs to different regions of the *gfp-mut2* promoter and open reading frame in *E.coli*, reduced 100-fold fluorescence expression on targeting regions (57). The other example is when dCas9 could repress β -galactosidase expression in *Streptococcus pneumoniae*. 14-fold reduction in β -galactosidase (*bga*) activity was observed after targeting *bgaA* promoter and open reading frame by sgRNAs and dCas9 (57). Similar study was done on the expression of inserted red fluorescent protein (mRFP) into the *E.coli* genome which resulted in silencing by dCas9 and sgRNA (61).

Besides these results, some points were achieved through these studies. First, different regulation systems of promoters may influence the CRISPRi gene repression (57). Second, if sgRNAs binds to coding sequence, effective gene silencing was observed than targeting the template strand. Additionally, CRISPRi is highly specific as no other genes besides the targeted gene showed reduction in expression (61). The other result was that the introduction of mismatches in the crRNA guide, could reduce the level of repression. Although the number of mismatches depends on the mode of dCas9 repression (57).

Some studies revealed that CRISPRi was also functional in eukaryotes such as mammalian cells. Qi et al. tested this system in the human embryonic kidney (HEK293) mammalian cells. dCas9 was fused to Nuclear localization sequence (NLS) and sgRNAs were expressed from the RNA polymerase III U6 promoter that targeted the nontemplate DNA strand of the EGFP-coding region. 46% repression was observed in EGFP expression (61). Similarly, another study on *Saccharomyces cerevisiae* revealed that dCas9:sgRNA could repress endogenous TEF1 locus (62).

Qi et al and gilbert et al studies showed that although dCas9 was functional in eukaryotes, its efficiency of repression was not significant. While the addition of transcriptional repressors to dCas9 resulted in improvement of gene repression (61, 62). For instance, the addition of a mammalian transcriptional repressor domain (Mxi1) that interacts with the histone deacetylase Sin3 homolog in yeast, increased repression of TEF1 locus in *S. cerevisiae* from 18-fold (dCas9 alone) to 53-fold. Similarly, dCas9-KRAB fusion protein resulted in 5-fold decrease in GFP in

HEK293 cells or greatly repressed transferrin receptor (CD71) and C-X-C chemokine receptor type 4 (CXCR4) in Hela cells (62).

Regarding gene repression in mammalian cells, nontemplate strand is more susceptible to repression than template strand. Additionally, CRISPRi efficiency is dependent on some factors such as precise targeting locus, the distance from the transcription start site, local chromatin state and stability of dCas9 and sgRNA in mammalian cells (61). Moreover, CRISPRi has this potential to silence multiple genes simultaneously without crosstalk (60, 61). By using different sgRNAs with distinct complementary regions to each gene and different fluorescence reporter system, each gene repressed separately with no effect on other genes (61). For example, transfection of sgRNAs which targeting CD71, CXCR4 or both and dCas9-KRAB into the Hela cells simultaneously resulted in simultaneous silencing of CD71 and CXCR4 (62).

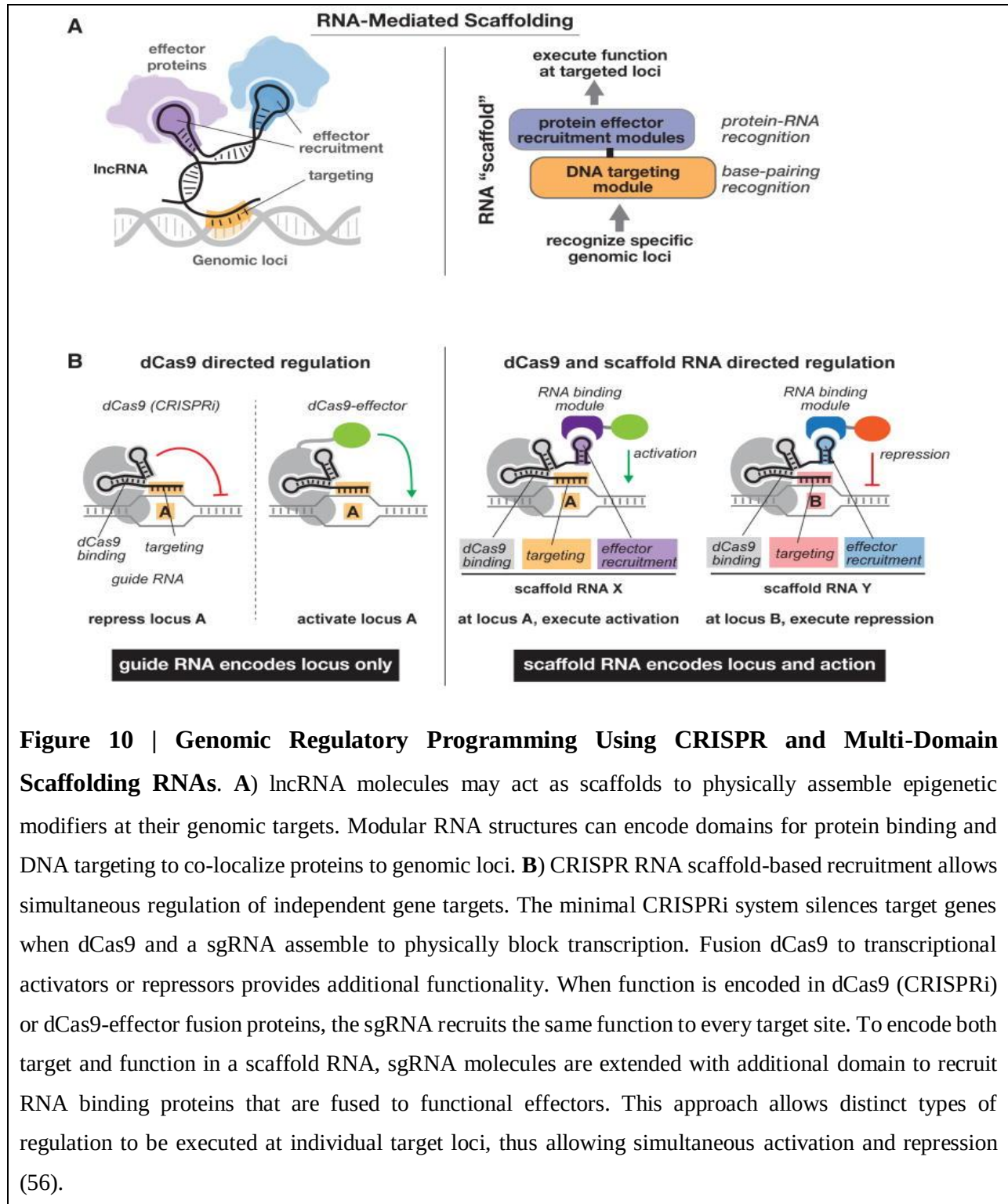
Activating gene expression is another potential of activating CRISPR (CRISPRa) which is done by the help of dCas9 and activators. Bikard et al. used CRISPRa to activate lacZ gene in *E.coli*. In their research, omega subunit (ω) of RNAP was fused to dCas9. Omega subunit can activate transcription by stabilizing the binding of RNAP to a promoter with an upstream lambda operator. As a result, when dCas9- ω fusion bound to optimal distance from the promoter, 2.8-fold activation was observed in LacZ expression (57). Similar attempts have been done in eukaryotic cells to show the efficiency of CRISPRa in eukaryotes (62, 63).

One of the efforts to test the ability of CRISPRa in endogenous gene activation in eukaryotes was done by Gilbert and his colleagues. dCas9-VP64 and dCas9-p65AD were co-transfected into a HEK293 reporter cell line. sgRNA was also constructed to target Gal4 UAS which was coexpressed with GFP. As a result, the expression of the reporter gene (Gal4 UAS-GFP) was effectively increased (62). Even though, CRISPRa was functional for in vivo studies. For example, in *Drosophila*, fusion of dCas9 with proper activators such as VP64 or p65 resulted in activation of different genes and expression predicted phenotypes for those genes: CRISPRa of the EGFR ligand vein in the wing led to excess vein tissue or activation of the FGF ligands *pyr* and *ths* caused similar defects in wing growth and patterning (63).

Although CRISPRa can cause activation in mammalian cells, this activation is usually moderate. Several strategies have been tested to enhance activating CRISPR in mammalian cells. One strategy is to recruit additional activators to the promoter by the help of several nonoverlapping

sgRNAs (59, 60). A tandem array of peptides fused to dCas9 which recruit multiple activators and could enhance activation from 2- to 50-fold (60). Another strategy is to fuse different activators together (59, 60, 63). For example, VP64, p65 and Rta (VPR) linked in tandem to the dCas9 and upregulated gene expression from 5- to 300-fold compared to a single dCas9-VP6 fusion (60).

Moreover, it is possible to activate and repress multiple genes within a single cell. In Zatalan et al. research the first step to achieve this multiple regulation was to turn the sgRNA in to a scaffold. This scaffold RNA (scRNA) had this ability to recruit different types of effectors. An RNA hairpin fused to the sgRNA and then a fusion of different effectors with RNA-binding proteins (RBP) bound to this hairpin structure. As a result, scRNA bound to target genes through DNA-base-pairing and regulated those genes (activating or repressing) through the additional RBP-effectors simultaneously. However, the number of fused hairpins and thus RNA stability is a limiting factor for gene regulation (56).



4.4. CRISPR-Cas9 as a means of DNA visualizing

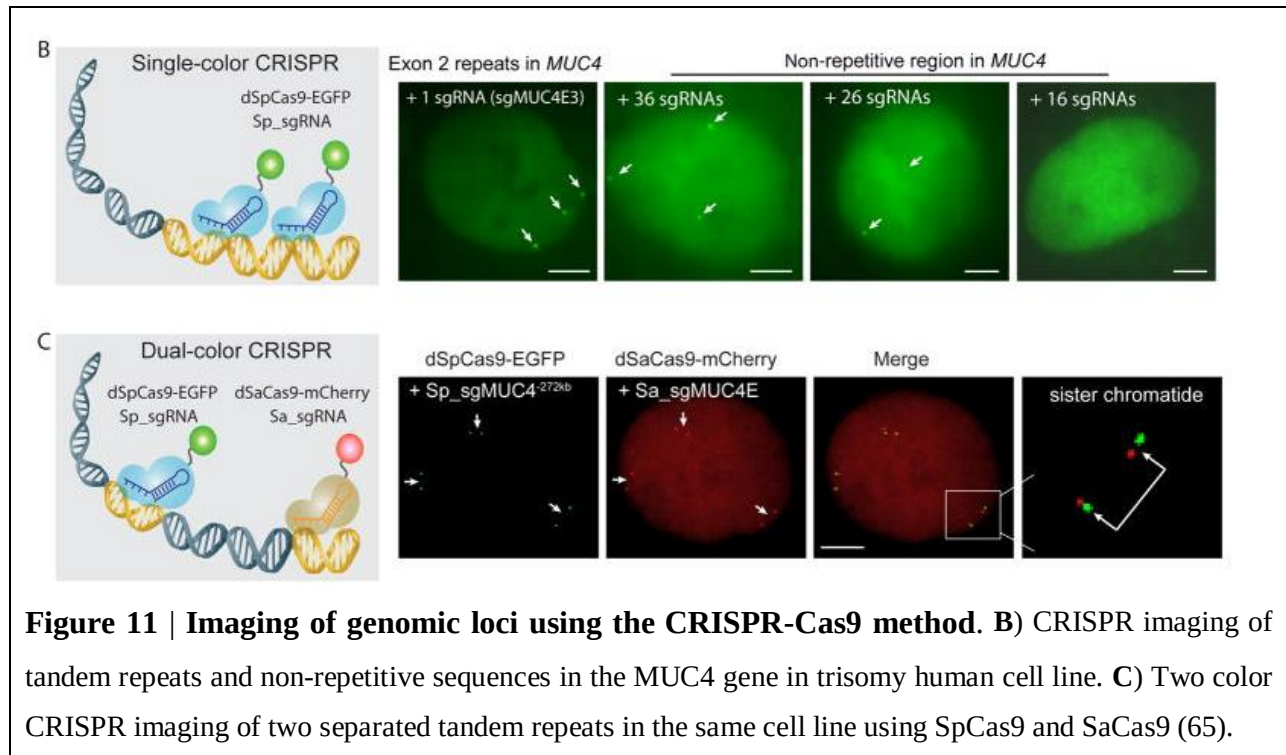
Another application of CRISPR-Cas9 is visualizing and imaging of genomic loci. The importance of imagination and defining the location of each locus is because of the relation between their organization and functions and mechanisms. Most of the visualizing methods have been based on fluorescent proteins that are capable on binding to DNA while these methods are restricted to imaging artificial repetitive sequences inserted into the genome or specialized genomic elements such as the telomeres or centromeres. For other methods such as fluorescence *in situ* hybridization (FISH), sample fixation and DNA denaturation is necessary while CRISPR-Cas9 passes these challenges (64, 65).

CRISPR-Cas9 has this ability to imagine endogenous genomic sequences (repetitive and non-repetitive) in mammalian and human cells specifically and dynamically (64, 65). sgRNA, dCas9 and fluorescent protein such as EGFP are three essential components for visualizing DNA by CRISPR. sgRNAs target the specific locus and bring dCas9 fluorescent protein fusion to the target locus. NLS also facilitate targeting in nucleus (64). In the other strategy, a fluorescent RNA-binding protein bind to the sgRNA by aptamer fusions and converting sgRNA into a scaffold RNA (scRNA). This scRNA has this ability to target locus and bring fluorescent color simultaneously (65).

Chen et al. research revealed the capability of CRISPR for detecting gene copy number in living cells. Chromosome 3 trisomy and aneuploidy of the Hela and UMUC3 cell lines were detected by CRISPR. Moreover, the movements of targeted gene, its replication behavior during late S phase and G2 phase and position in the nucleus and chromosome conformation during mitosis were monitored by CRISPR-Cas9 (64). Some genes expressed and repressed periodically in cells such as cyclins. Regarding these type of genes, CRISPR-Cas9 can show the dynamics and periodic changes of these genes and so can be useful in regulating gene expression (65).

Moreover, there are some studies to improve the efficiency of CRISPR-Cas9 in visualizing DNA. For example, different variants of dCas9 were fused to different fluorescent proteins and so could image different genomic sequences with different colors. Additionally, by using multiple sgRNAs, two genes were labelled and imaged simultaneously (64). Multicoloring method is so useful for

monitoring non-repetitive sequences and determining the intracellular distance between loci on different chromosomes or the spatial resolution between two loci on the same chromosome (65).



4.5. More Advantages of CRISPR-Cas9 in Comparison with Previous Systems

The other crucial component of CRISPR-Cas9 is a sgRNA library which is more advantageous than previous systems. As mentioned earlier, making a sgRNA library (37, 66) by different techniques is faster, cheaper and more efficient than protein synthesis in TALENs. In addition, even in a large scale screen, a large proportion of sgRNAs successfully generate mutations at their target sites. They could separate true and false positives on a gene level perfectly (15, 67).

With respect to previous technologies, there are three particular advantages over ZFNs and TALENs. 1). Simplicity: the 20-nt guide sequence within a sgRNA can be easily programmed by PCR or *de novo* synthesis to achieve new DNA targeting specificity. 2) multiplexability: multiple sgRNAs can work with the same Cas9 protein on many different target sites simultaneously, and

3) Insensitivity to target DNA methylation. This feature enables the CRISPR-Cas9 system to edit on the genomic regions under heavy epigenetic regulation (66).

Table 1 shows comparisons between the advantages of CRISPR-Cas9 and TALENs from various aspects. A brief explanations and reasons for each factors are included as well.

Table 1: Comparison between CRISPR-Cas9 and TALENs

Factors	CRISPR-Cas9	TALENs	Explanations
Gene editing	Powerful and Multiplexable	Complex and singular gene editing	- TALEN nuclease recognition site needs T (thymidine) before the 5'- end of the target sequence and targets distinct loci (67) while in CRISPR-Cas9 system By using sgRNAs targeting different genomic loci and packing them with the same dCas9 protein in one vector could perform different functions at different target sites, therefore allowing for functional multiplexing using one CRISPR-Cas9 system (1, 67, 69). - Coinjection of single-stranded oligos can introduce designed point mutations into two target genes in one step, allowing for multiplexed gene editing in a strictly controlled manner (69).

Time	Time-saving	Time-consuming	-In TALENs, two proteins must be designed and engineered for each target (70, 71) but in CRISPR method all components (Cas9, sgRNA, and oligos) transferred simultaneously and therefor resulted in multiple genome editing in less time (1, 67, 69).
Price	Cheap	More expensive than CRISPR	-Targeting distinct loci with TALENs requires the design and synthesis of a new custom TALE array for each locus, which is more costly (70).
Design	Easy	Hard	-A growing number of software tools have become available for designing guide RNA and predicting off-target profiles for CRISPR systems. These tools allow specification of experimental goals by users for different desired modifications (e.g., insertion of tags, disruption of protein domains, etc.) or allow the use of gene architecture annotation to assist guide RNA designs (1, 67).

Genome editing in animals	Yes	Yes	-Injection of CRISPR/Cas9 components (Cas9 messenger RNA or protein; sgRNA; HDR template) into the zygote or early stage embryo allows modifying the genome in all cells of the organism, including the germline. Thus, this approach results in permanent changes that can be passed on to subsequent generations, offering the possibility to eliminate a genetic disease from an entire family (67).
Genome editing in plants	Yes	No	CRISPR has potential roles in creating variation in germoplasm and applying this interference pathway in crop improvement. Also the availability of CRISPR/CAS9 system facilitates research in crops that lack genetic resources (55).
Specificity and Precision	High	Lower than CRISPR	-The protospacer-associated motif (PAM) NGG sequence, which is indispensable for cleavage of DNA by Cas9, distinguishes self from non-

			self DNA and 20 nt guide sequence located at the 5' end of the sgRNA determine targeting specificity (72). -HDR repair system in CRISPR method allows rewriting of DNA sequence and generation of new gene or protein sequences (72).
--	--	--	---

*(Because ZFNs are often difficult to engineer, we have chosen to compare Cas9 with TALEN only.)

Chapter 5: Other Accomplishments

Because of the mentioned potentials of CRISPR-Cas9 system, innovative and valuable achievements have been reached. In this chapter, a summary of some of the notable work done using this system will be presented.

5.1. CRIPSR/Cas9 in *T. cruzi* Genome Screens

Trypanosome cruzi (*T. cruzi*) is a parasite that causes human Chagas disease. In the acute phase of this infection, parasite is in the host blood stream and other tissues. This infection promotes severe inflammatory response especially in heart muscle. Infection diagnostics and treatment options are poor and limited. In addition, suitable vaccines do not exist (73).

Adapting CRISPR-Cas9 for use in *T. cruzi* shows another advantage of this system. Manipulation of the *T. cruzi* genome and the processes to achieve modifications are not rapid, easy, or routine. Because of the low rate of homologous recombination and the relatively high resistance of *T. cruzi* to drug selection, production of null mutants of single-copy genes requires a minimum of several months and the use of multiple drug resistance genes. The latter limitation is a particular problem in *T. cruzi*, which has over 100 gene families with 4 or more members and several families that contain hundreds of genes (73).

By contrast, CRISPR-Cas9 provides this opportunity to quickly and efficiently disrupt endogenous single-copy and multi-copy genes, as well as exogenous genes. Because of the high rate of gene mutation when using CRISPR-Cas9 in *T. cruzi*, routinely 60 to 70% double-/multiallelic mutations, the impact of gene disruption can be observed within days when appropriate assays are available (e.g., fluorescence for GFP, enzyme activity for HAL, and fatty acid uptake for FATP) instead of months, as in the case of conventional knockouts (73).

5.2. Identifying Essential Genes for West Nile virus-induced Cell Death

Another knockout screening method based on CRISPR-Cas9 was performed on cultured cells to identify essential genes for West Nile virus (WNV)-induced cell death. WNV infection causes massive neuronal cell death, mortality and morbidity. Their strategy was to disrupt target genes by a sgRNA library. Cells harboring the sgRNA library were transfected with a plasmid that transiently expresses Cas9 (34)

Three independent experiments were done and multiple sgRNAs with a high number of reads were seen for three genes, EMC2, EMC3, and SEL1L. The investigators performed a second round of screening by using a sub-library of sgRNAs that targeted the identified genes in the first round with more than five sgRNAs in three experiments. This round was to distinguish essential WNV-induced cell death genes from background noise to improve the specificity and sensitivity of the test (34).

As a result, seven genes were identified with the highest number of sgRNAs in each experiment. These genes belong to the READ pathway which appear to be essential for WNV infection to cause cell death. In vitro studies in 293FT cells (a variant of human embryonic kidney cells), neuronal cells and Hela cells revealed that although viral replication continues in the absence of those seven genes, cells survive (34).

5.3. Therapeutic Approach of CRISPR-Cas9 for Individual with Duchenne muscular dystrophy (DMD)

Another application of the CRISPR-Cas9 tool is to modulate gene expression for therapeutic use. For instance, in one study, researchers aimed to upregulate utrophin in skeletal muscle cells of an individual with Duchenne Muscular Dystrophy (DMD). DMD is an X-linked recessive neuromuscular disorder associated with muscle degeneration because of mutations in DMD and absence of the protein dystrophin. In their approach, myoblasts of a DMD-affected individual were transfected with a dead Cas9 (dCas9) and several sgRNAs to target either the UTRN A or UTRN B promoters. This led to upregulation of utrophin higher than basal amounts and restored the amount of B-dystroglycan in muscle cells of individuals with DMD in vitro (72, 74).

Furthermore, they provided evidence that their designed sgRNA could successfully target two promoters. Here a strategy was developed that CRISPR-Cas9 system could remove duplicated regions within the genome thus restoring gene function. Using the least amount of CRISPR-Cas9 components (one sgRNA) in their strategy was another interesting point that will be critical for the development of further therapeutic applications. Although it has been unclear whether this type of genetic correction would restore a fully functional gene, this system opens up new treatment strategies for individuals affected by DMD duplications (74).

5.4. Treating Tyrosinaemia Type 1 with CRISPR-Cas9

CRISPR-Cas9 system has also been used for treating diseases like tyrosinaemia type 1. In a mouse model, investigators deleted the hydroxyphenylpyruvate dioxygenase (Hpd) gene in hepatocytes, thereby successfully rerouting tyrosine catabolism, resulting in healthy and asymptomatic mice. Pharmacological treatment can be quite effective for a period of time with numerous drawbacks. For example, it may not entirely inhibit target enzyme or upregulate the target. Chemical interactions of drugs or off-target effects are other examples. Thus CRISPR-Cas9 can be a better remedy than pharmacological treatment because it has not shown those mentioned problems (75).

Chapter 6: Challenges with CRISPR-Cas9

Despite the potentials of CRISPR-Cas9 in genome editing, there are some important issues that need to be addressed, such as off-target mutations, PAM dependence, and gRNA production and delivery methods of CRISPR-Cas9 (51).

6.1. Off-target Effects

Off-target mutations are one major concern associated with CRISPR-Cas9-mediated genome editing. Large genomes often contain multiple DNA sequences that are identical or homologous to target DNA sequences. Besides target DNA sequences, CRISPR-Cas9 could also cleave these identical or highly homologous DNA sequences, which can lead to off-target mutations at undesired sites. Compared with ZFNs and TALENs, CRISPR-Cas9 presents relative high risk of off-target mutations in human cells. Even minor off-target mutations could potentially result in cells death, transformation, or other unwanted phenotypes (51).

6.2. PAM Dependence

The specificity of CRISPR-Cas9 requires a 2–5 nt PAM sequence located immediately downstream of the target sequence. The identified PAM sequences vary among different Cas9 orthologs. The PAM-dependent manner of CRISPR-Cas9-mediated DNA cleavage constrains the frequencies of targetable sites in genomes. For instance, it is possible to find a target site per 8 nt for NGG PAM and NAG PAM, while per 32 and 256 nt for NGGNG PAM and NNAGAAW PAM (51).

6.3. sgRNA Production

sgRNA production is another important issue for CRISPR-Cas9. Due to extensive posttranscriptional processing and modification of mRNA transcribed by RNA polymerase II, it is currently difficult to apply RNA polymerase II for gRNA production. RNA polymerase III, U3 and U6 sgRNA promoters are currently used to produce gRNA *in vivo*. However, U3 and U6 sgRNA genes are ubiquitously expressed housekeeping genes, which cannot be used to generate tissue- and cell-specific gRNA. The lack of commercially available RNA polymerase III also limits the application of U3- and U6-based gRNA production (51).

6.4. Delivery Methods

Questions also remain regarding the delivery methods of CRISPR-Cas9 into organisms (34). DNA and RNA injection-based techniques are used for CRISPR-Cas9 delivery, such as injection of plasmids expressing Cas9 and gRNA and injection of CRISPR components as RNA. In most cases, Cas9 variant and each sgRNA are encoded on independent plasmids and mixed for co-transfection. However, co-transfection of multiple plasmids leads to variable expression levels in each cell due to differences in copy number (51).

In addition, gene activation following transfection is transient due to dilution of plasmid DNA over time. This dilution could occur because of plasmid replication or degradation. There are some nucleases in cytosol of cells which could degrade external plasmids. Additionally, many cell types are not easily transfected, and transient gene expression may not be sufficient for achieving the desired effect (41). The efficiencies of delivery methods depend on the types of target cells and tissues (51).

Chapter 7: Conclusion and discussion

As we have described, there are multiple genome-editing technologies that are currently used to induce changes in the genome (2). These methods have provided a revolutionary step in advances toward treatments of genetic diseases. Compared to the previous genetic engineering systems like ZFNs and TALENs, the newly developed CRISPR-Cas9 system has the potential to become a more reliable and useful genome editing tool (51).

Two important and essential components of this system are Cas9 and sgRNA. These components work in parallel to detect target sequences in the genome followed by the induction of a double strand break in that region. By providing a pooled library of sgRNAs, new variations of Cas9, and optimizing the technology, the specificity and sensitivity of CRISPR-Cas9 technology will likely increase (2,5).

One of the applications of CRISPR-Cas9 technology was in finding positive and negative regulatory genes for TNF- α induction after exposure to LPS. Using this technology, investigators found new genes that were previously not known to play a role in the TNF signaling pathway (15). Identifying essential genes for WNV-induced cell death which belong to the READ pathway was another valuable accomplishment that was achieved through the CRISPR-Cas9 system (35). Other CRISPR based accomplishments include the mouse models that show efficacy in the treatment of diseases like tyrosinaemia type 1(75) and upregulation of utrophin in muscle cells of individuals with DMD (72, 74).

Table 2: Advantages and Disadvantages of CRISPR-Cas9

Advantages	Disadvantages
Cheap-method	Off target mutations
Easy to design and perform	PAM dependence
Simultaneous multiple genetic modification	gRNA production
Facilitate generation of mutations	Delivery methods

Overall, further studies will be necessary to improve the performance and efficacy of the CRISPR-Cas9 system. For instance, Parnas et al., demonstrated the validity of the CRISPR-Cas9 system by

performing three independent screens, with different amount of sgRNAs per gene to confirm the specificity and sensitivity of the CRISPR-Cas9 method (15). Their results suggested that the CRISPR-Cas9 system's potential depends on the number of sgRNAs and infected cells.

One of the main challenges associated with the CRISPR-Cas9 system is off-target effects (51). Although it seems that this challenge did not substantially affect the results obtained thus far (15, 35, 71), reducing off-target effects could improve the utility of this system. For example, if one could synthesize new variants of Cas9 that do not need PAM sequences or recognize more specific sequences as PAM, this would represent an advance in the system. Another potential advance would be synthesizing larger sgRNAs or those with complementary sequences of noncoding region in their beginning and ending points, resulting in less off-target cleavage or cleavage on noncoding regions.

References:

- 1- Nemudryi, Valetdinova, Medvedev, Zakian. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Nature*. **6** (3): 19-40 (2014).
- 2- Wang, et al. Generation of gene-modified goats targeting *MSTN* and *FGF5* via zygote injection of CRISPR/Cas9 system. *Sci Rep*.**10**: 13787 (2015).
- 3- Gaj, Gersbach, Barbas. ZFN, TALEN, and CRISPR/Cas-based Methods for genome engineering. *Trends in Biotechnology*. **31**: 397-405 (2013).
- 4- Christian, et al. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *GENETICS*. **186** : 757-761 (2010)
- 5- Wilkinson and Wiedenheft. ACRISPR method for genome engineering. *F1000Prime Rep*. **6**:3 (2014).
- 6- Shalem, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*. **3**:84-7 (2014).
- 7- Mahfouz, Piatek, and Neal Stewart Jr. Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. *Plant Biotechnol*. **12**: 1006-1014 (2014).
- 8- Kim and Soo Kim. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*. **15**: 321-334 (2014).
- 9- Cermak, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acid Research*. **39**(12): e82 (2011).
- 10- Weber, Gruetzner, Werner, Engler, and Marillonnet. Assembly of Designer TAL Effectors by Golden Gate Cloning. *PLOS One*. **6**(5): e19722 (2011).
- 11- Lamb, Mercer, Barbas. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res*. **41**: 9779-9785 (2013).
- 12- Liang, Zhang, Lou, Yu. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Scientific Reports*. **6**: 21451 (2016).
- 13- Nowak, Lawson, Zerez, Bleris. Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res*. **44**: 9555-9564 (2016).
- 14- Ran, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. **8**: 2281-2308 (2013).
- 15- Parnas, et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell*. **162**:675-686 (2015)

- 16- Da Silva correia, Ulevitch RJ. MD-2 and TLR4 N-linked glycosylations are important for a functional lipopolysaccharide receptor. *J Biol Chem.* 18:1845-54 (2002).
- 17- Carballo, Lai, Blackshear. Feedback Inhibition of Macrophage Tumor Necrosis Factor- α Production by Tristetraprolin. *Science.* **14**: 1001-5 (1998).
- 18- Zhang, Yao. The Physiological Role of Tumor Necrosis Factor in Human Immunity and Its Potential Implications in Spinal Manipulative therapy: A Narrative Literature Review. *J.chiopr Med.* **15**: 190-6 (2016)
- 19- Parameswaran and patial. Tumor Necrosis Factor- α signaling in Macrophages. *Crit Rev Eukaryot Gene Expr.* **20**: 87-103 (2010).
- 20- Aringer and smolen. SLE-complex cytokine effects in a complex autoimmune disease: tumor necrosis factor in systemic lupus erythematosus. *Arthritis Res Ther.* **5**: 172-177 (2003).
- 21- Dagvadorj, et al. Interleukin-10 inhibits tumor necrosis factor- α production in lipopolysaccharide-stimulated RAW 264.7 cells through reduced MyD88 expression. *Innate Immun.* **14**: 109-15 (2008).
- 22- McDermott, AJ O'Neill. Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4. *Immunology.* **113**: 153-162 (2004).
- 23- Bruggen, Nijenhuis, Raaij, Verhoef, Asbeck. Lipopolysaccharide-Induced Tumor Necrosis Factor Alpha Production by Human Monocytes Involves the Raf-1/MEK1- MEK2/ERK1-ERK2 Pathway. *Infect Immun.* **67**:3824-3829 (1999).
- 24- Xanthoulea, et al. Tumor necrosis factor (TNF) receptor shedding controls thresholds of innate immune activation that balance opposing TNF functions in infectious and inflammatory diseases. *J Exp Med.* **200**:367-76 (2004).
- 25- DCeci, et al. TNF- α Induction by LPS Is Regulated Posttranscriptionally via a TP12/ERK-Dependent Pathway. *Cell.* **103**: 1071-1083 (2000).
- 26- O'Brien et al. Tumor Necrosis Factor Alpha Receptor I Is Important for Survival from Streptococcus pneumonia Infections. *Infect Immun.* **67**: 595-601 (1999).
- 27- Wanidworanun C, Strober W. Predominant role of tumor necrosis factor- α in human monocyte IL-10 synthesis. *J Immunol.* **151**: 6853-61 (1993).
- 28- Chen, J. Oppenheim. Contrasting effects of TNF and anti-TNF on the activation of effector T cells and regulatory T cells in autoimmunity. *FEBS Lett.* **585**: 3611-3618 (2011).
- 29- Diamond, Khameneh, Brough, and Mortellaro. Novel perspective on non-canonical inflammasome activation. *Immunotargets Ther.* **4**: 131-141 (2015).

- 30- Gamucci, et al. Detection of Fluorescent Nanoparticle Interactions with Primary Immune Cell Subpopulations by Flow Cytometry. *J Vis Exp.* **85**: 51345 (2014).
- 31- Herman, Cermak, Voytas, Pelczar. Mouse genome engineering using designer nucleases. *J Vis Exp.* **2** (2014).
- 32- Carroall, et al. A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proc Natl Acad Sci U S A.* **2**: 338-43 (2016).
- 33- Cho, Haruyama, Kulkarni. Generation of Transgenic Mice. *Curr Protoc Cell Biol.* Unit 19.11 (2009).
- 34- Ma, et al. A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. *Cell Rep.* **12**(4): 673-83 (2015).
- 35- Barrangou and May. Unraveling the potential of CRISPR-Cas9 for gene therapy. *Expert Opinion on Biological Therapy.* **15**: 311-314 (2015).
- 36- Carrington, Varshney, Burgess, Sood. CRISPR-STAT: an easy and reliable PCR-based method to evaluate target-specific sgRNA activity. *Nucleic Acids Res.* **15**: e157 (2015).
- 37- Schmidt, L. Schmid-Burg, Homung. Synthesis of an arrayed sgRNA library targeting the human genome. *Sci Rep*, **5**: 14987 (2015).
- 38- Cheng, et al. A Molecular Chipper technology for CRISPR sgRNA library generation and functional mapping of noncoding regions. *Nature communications.* **7**: 11178 (2016).
- 39- Benskey MJ, Manfredsson FP. Lentivirus Production and Purification. *Method Mol Biol.* **1382**: 107-14 (2016).
- 40- Fu, Sander, Reyon, Cascio, Joung. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol.* **32**: 279-284 (2014).
- 41- M. Kabadi, Ousterout, Hilton, A. Gersbach. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* **42**: e147 (2014).
- 42- E. Heyer, Ozadam, Ricci, Cenik, and J. Moore. An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Res.* **43**: e2 (2015).
- 43- Noman AS, et al. Thalidomide inhibits lipopolysaccharide-induced tumor necrosis factor- α production via down regulation of MyD88 expression. *Innate Immune.* **15**: 33-41 (2009).
- 44- Singh. microRNAs Databases: Developmental Methodologies, Structural and Functional Annotations. *Interdiscip Sci* (2016).
- 45- Lubos, et al. Glutathione peroxidase-1 modulates lipopolysaccharide-induced adhesion molecule expression in endothelial cells by altering CD14 expression. *FASEB J.* **24**: 2525-2532 (2010)
- 46- Harada, et al. Eukaryotic Oligosaccharyltransferase Generates Free Oligosaccharides during N-Glycosylation. *J Biol Chem.* **8**: 32673-32684 (2013).

- 47- Youn MY, et al. hCTR9, a component of Paf1 complex, participates in the transcription of interleukin 6-responsive genes through regulation of STAT3-DNA interactions. *J Biol Chem.* **30**: 34727-34 (2007).
- 48- Chaudhary, Deb, Moniaux, Ponnusamy, Batra. Human RNA polymerase II-associated factor complex: dysregulation in cancer. *Oncogene.* **26**: 7499-507 (2007).
- 49- Lu, et al. Macrophage inflammatory protein-1 α promotes the growth of acute myeloid leukemia cells. *Zhongguo Shi yan Xue Ye Xue Za Zhi.* **23**: 306-11 (2015).
- 50- Yu, et al. Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell Stem Cell.* **16**: 142-147 (2015).
- 51- Zhang, Wen, Guo. CRISPR/Cas9 for genome editing: progress, implications, and challenges. *Hum Mol Genet.* **15**:R40-6 (2014).
- 52- Ma , Zhang , Huang . Genome modification by CRISPR/Cas9. *FEBS J.* 281: 5186-93 (2014).
- 53- Bothner, et al. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun.* **8**: 13905 (2017).
- 54- Zeng, et al. The initiation propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Research.* **46**: 350-361 (2018).
- 55- Tsai, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol.* **32**: 569-576 (2014).
- 56- Zalatan, et al. Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffold. *Cell.* **160**(0): 339-350 (2015).
- 57- Bikard, et al. Programmable repression and activation of bacterial gene expression using engineered CRISPR-Cas system. *Nucleic Acids Research.* **41**: 7429-7437 (2013).
- 58- Park, Dempewolf, Zhang, and Yu Wang. RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in Arabidopsis. *PLOS One.* **12**(6): e0179410 (2017).
- 59- Sayman, et al. Potent Targeted Activation of Latent HIV-1 Using the CRISPR/dCas9 Activator Complex. *Cell.* **24**: 488-498 (2016).
- 60- La Russa and Qi. The New State of the Art: Cas9 for Gene Activation and Repression. *Cell. Biol.* **35**: 3800-3809 (2015).
- 61- Qi, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell.* **15**(5): 1173-1183 (2013).
- 62- Gilbert, et al. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell.* **154**(2): 442-451 (2013).

- 63- Ewen-Campen, et al. Optimized strategy for in vivo Cas9-activation in *Drosophila*. *Proc Natl Acad Sci U S A*. **114**(35): 9409-9414 (2017).
- 64- Chen, et al. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*. **155**(7): 1479-1491 (2013).
- 65- Chen, Guan, and Huang. Imaging Specific Genomic DNA in Living Cells. *Annu Rev Biophys*. **45**: 1-23 (2016).
- 66- Marie, La Russa, Lei S. The New State of the Art: Cas9 for Gene Activation and Repression. *Mol Cell Biol*. **35**: 3800-3809 (2015).
- 67- Wiles, Qin, Cheng, and Wang. CRISPR-Cas- mediated genome editing and guide RNA design. *Mamm Genome*. **26**: 501-510 (2015).
- 68- Zhang, Li, Feng Li. Targeted Gene Manipulation in Plants Using the CRISPR/Cas Technology. *Journal of Genetics and Genomics*. **43**:251-262 (2016).
- 69- Strauß and Lahaye. Zinc Fingers, TAL Effectors, or Cas9-Based DNA Binding Proteins: What's Best for Targeting Desired Genome Loci? *Molecular Plant*. **6**: 1384-1387 (2013).
- 70- Eid and Mahfouz. Genome editing: the road of CRISPR/Cas9 from bench to clinic. *Experimental & Molecular Medicine*. **48**: e265 (2016).
- 71- Hsu, Lander, and Zhag. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*. **157**: 1262-1278 (2014).
- 72- Savić and Schwank. Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research*. **168**:15-21 (2016)
- 73- Peng, Kurup, Yao, Minning, Tarleton. CRISPR-Cas9-Mediated Single-Gene and Gene Family Disruption in *Trypanosoma cruzi*. *mBio*. **6**: 1-14 (2014).
- 74- Wojtal, et al. Spell Checking Nature: Versatility of CRISPR/Cas9 for Developing Treatments for Inherited Disorders. *Am J Hum Genet*. **98** (1): 90-101(2016).
- 75- Pankowicz, et al. Reprogramming metabolic pathways in vivo with CRISPR/Cas9 genome editing to treat hereditary tyrosinaemia. *Nat Commun*. **7**:12642 (2016).