이학박사학위논문

Versatile application of the CRISPR-Cas system to various organisms

다양한 생명체에서의 크리스퍼-카스 시스템 적용에 대한 연구

2018년 2월

서울대학교 대학원

생명과학부

유 지 현

Abstract

Versatile application of the CRISPR-Cas system to various organisms

Jihyeon Yu

School of Biological Science

The Graduate School

Seoul National University

CRISPR-Cas is an adaptive immune system in bacteria and archaea. It specifically finds target through a complementary DNA-RNA hybridization and removes target DNA or RNA. A genome engineering has advanced explosively due to an introduction of CRISPR-Cas, moreover numerous studies have been accomplishing with CRISPR-Cas in most of scientific fields. Studies in below were a few instances of various applications using CRISPR-Cas.

The first study developed a cancer diagnosis method in a liquid biopsy using CRISPR-Cas. CUT (CRISPR-mediated, Ultrasensitive detection of Target DNA)-PCR can detect the extremely small amounts of tumor DNA fragments among the much more abundant wild-type DNA fragments by specifically eliminating the wild-type sequences with CRISPR endonucleases and enriching tumor DNA. I

computed that by using various orthologonal CRISPR endonucleases, the CUT-PCR method would be applicable to 80% of known cancer-linked substitution mutations. I further verified that CUT-PCR together with targeted deep sequencing enables detection of a broad range of oncogenes with high sensitivity

Second study showed a generation of *Crltd* mutants by DNA-free ribonucleoprotein delivery in microalgae, *Chlamydomonas reinhardtii*. *Crltd* mutants which were generated using CRISPR-Cas had a pale-yellow coloration and reduced chlorophyll and PSI complex contents than wild type. As a results, it was supposed that CrLTD involved in a trafficking of LHCP to thylakoid membranes in a chloroplast. An analysis of off-target mutations in *Crltd1* mutant confirmed that there were no unwanted mutations by an off-targeting. This study also optimized the method of the ribonucleoprotein delivery in *C. reinhardtii*, thus made much easy the generation of mutants by CRISPR-Cas.

Finally, I tried a targeted DNA methylation using inactive Cas9 protein and DME protein which is a plant-specific DNA demethylase. A catalytic domain of DME fused to inactive Cas9, it was introduced to human cells and plants to induce a DNA demethylation at specific regions. In human cells, no significant changes were observed in the target region, *KLF4* CpG island. In plant, *FWA* promoter region were targeted for generating *fwa* epi-mutants. Some of transgenic plants showed late flowering and vegetative axillary meristem, but DNA methylation levels of these plants were comparable to wild type. In addition, phenotypes of

transgenic plants were not inherited in a next generation. It proved unsuccessful to develop an epi-genetic tool using dCas9-DME as present, it is expected that the tool could be improved by further experiments for an optimization. The epi-genome engineering has an advantage of changing phenotypes without DNA information changes. Therefore, dCas9-DME seems to have a potential as a tool for the epi-genome engineering.

Keyword : CRISPR-Cas, genome engineering, cancer diagnosis, ctDNA, ribonucleoprotein, DNA demethylation

Student Number : 2007-20351

Table of Contents

List of Figures

Figure 1 was contributed by Sangsu Bae and Gue-Ho Hwang.

Figure 2, 5, 6, 7, 8, 9 and 10 were contributed by Seung Hwan Lee, Sojung Kim, Heon Seok Kim, Sunghyeok Ye and Kyoungmi Kim.

Figure 12 and 13 were contributed by Jooyeon Jeong.

List of Tables

Table 2-4 were contributed by Sangsu Bae and Gue-Ho Hwang.

Table 11 was contributed by Jin-sup Park.

List of Abbreviations

CRISPR	Clustered regularly Interspaced Short Palindromic Repeats
Cas	CRISPR-associated
DSB	Double strand break
HDR	Homology Directed Repair
Indel	Insertions and Deletions
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
MMEJ	microhomology-mediated end joining
gRNA	guide RNA
sgRNA	single-guide RNA
crRNA	CRISPR RNA
tracrRNA	trans-encoded crRNA
PAM	Protospacer Adjacent Motif
ZFN	zinc finger nuclease
TALEN	transcription-activator-like effector nuclease
COSMIC	Catalogue of Somatic Mutations in Cancer
ctDNA	circulating tumor DNA
cfDNA	cell-free DNA
CUT-PCR	CRISPR-mediated, Ultrasensitive detection of Target DNA-PCR
CRC	colorectal cancer
RNP	ribonucleoprotein

LHCP	light-harvesting chlorophyll a/b-binding protein
GMO	Genetically Modified Organism
5mC	5-methylcytosine
5hmC	5-hydroymethylcytosine
5fC	5-formylcytosine
5caC	5-carboxylcytosine
BER	base excision repair

Chapter I. Background

The genome editing is a manipulation of genome, a process of modifying DNA of organisms in a targeted manner. Recent advances in genome editing tools expand an area of the genome editing from a DNA modification to epi-genome modification and RNA modification (Hsu et al. 2014, Hilton et al. 2015, Liu et al. 2016, Vojta et al. 2016, Cox et al. 2017).

The early efforts to modify the eukaryotic genome were an introduction of donor DNA by homologous recombination (HR). The efficiency of HR in eukaryotic cell is extremely low, but it is greatly increased by a targeted DNA double-strand breaks (DSBs)(Rudin et al. 1988).

Hereafter programmable nucleases had been focused on the genome editing tools. ZFNs (Zinc Finger Nucleases) is the first generation of genome editing tools. A Combination of zinc finger modules fused to the FokI nuclease and paired ZFNs cleaves target DNA specifically (Kim et al. 1996). However ZFNs had a low activity and a problem of its cytotoxicity (Kim 2016). In a second-generation, DNA binding module was replaced by *Xanthomonas*-drived transcription activator-like effector (TALE) protein (Boch et al. 2009, Moscou et al. 2009). TALE has also modular structure in DNA binding domain, but TALE repeat recognizes a single base. Even though TALEN has higher activity and lower toxicity than ZFNs, an emergence of CRISPR-Cas makes it an old-fashioned tool.

CRISPR (Clustered, regularly interspaced short palindromic repeat) system is bacterial adaptive immune system for an antiviral

defence (Marraffini 2015). CRISPR loci are clusters of short sequences (protospacer) that are separated by partially palindromic, repetitive sequence (spacer) (Makarova et al. 2006). A protospacer is originated from foreign DNA like phage or plasmid DNA. Foreign DNA is processed to short sequence and integrated between spacer sequences in a CRISPR array by a Cas1-Cas2 complex (Nuñez et al. 2014). A mechanism of immunity by CRISPR is quite different to a type of CRISPR-Cas9. In case of type II Cas9, Pre-crRNA (precursor crRNA) is transcribed from CRISPR array, then process by RNase III (Deltcheva et al. 2011). A processed crRNA and tracrRNA (trans-encoded crRNA) are associated to an effector protein Cas9. Cas9-crRNA complex binds to target foreign DNA which has PAM (protospacer-adjacent motif) and cleaves DNA double strand by its own nuclease activity (Anders et al. 2014, Jinek et al. 2014, Nishimasu et al. 2014). CRISPR-Cas specifically targets and cleaves foreign DNA, thus a duplication of phage and plasmid are suppressed in a bacterial cell (Garneau et al. 2010).

Many bacteria and archaea have CRISPR-Cas locus in their genome. In analyzed strains, 87% of archaea and 50 % of bacteria have CRISPR-Cas system (Makarova et al. 2015). CRISPR-Cas systems classified to two classes (Koonin et al. 2017). Class 1 has multi-subunit effector complexes and Class 2 has single-protein effector. Class2 CRISPR divided to three categories by structures of effector proteins and target molecules. Class 2-type II CRISPR includes widely-used SpCas9 originated from *Streptococcus pyogenes*. Cas12a/Cpf1 which belongs to Class 2-type V CRISPR

cleaves DNA with staggered end. RNA-targeting Cas13a/c2c2 is classified Class 2-type VI.

Recently the mechanism that CRISPR-Cas cleaves target DNA has been studied enthusiastically. I would explain briefly a mechanism of CRISPR-Cas based on type II CRISPR-Cas9. TracrRNA and crRNA associated with Cas9 protein then Cas9-gRNA complex searches proper target sequences. A PI motif in Cas9 protein recognizes PAM sequence in a genome, unwinding of DNA double strand occurs (Anders et al. 2014, Jinek et al. 2014, Nishimasu et al. 2014). If crRNA matches to target strand and RNA-DNA hybridization is stabilized, conformational change of HNH domain occurs (Dagdas et al. 2017). It leads to Cas9 complex to active state. Target strand of DNA is cut by HNH domain and non-target strand is cut by RuvC domain of Cas9.

DNA DSB by programmed nucleases activates an intrinsic DNA repair system (Ceccaldi et al. 2015). According to cell environment, DNA repair proceeds to three pathways. NHEJ/c-NHEJ (non-homologous end joining) pathway and MMEJ/a-NHEJ (microhomology-mediated end joining) are error-prone repair and can form a small insertion or deletion (indel). Thus, this pathway can induce gene knock-out resulting from a frame-shift in coding sequences. HDR (homology directed repair) rarely occurs and requires a donor template like as sister chromatid or plasmid DNA. A precise genome editing for the gene correction is feasible through a HDR pathway.

Ever since CRISPR-Cas was introduced as a genome editing

tool (Jinek et al. 2012, Cho et al. 2013, Cong et al. 2013, Jinek et al. 2013, Mali et al. 2013), applications of CRISPR-Cas have increased explosively (Kim 2016). A major advantage of CRISPR-Cas is very easy to use. It requires only a Cas9 protein and a target-specific gRNA. Unlike ZFNs and TALENs which need a pair of new proteins to new target site, CRISPR-Cas only needs a new gRNA that can be easily cloned to plasmid or synthesized *in vitro*.

I would represent applications using CRISPR-Cas to various organism and subjects. They could show an infinite probability of CRISPR-Cas as a biological tool.

Chapter II. Cancer diagnosis using CRISPR-Cas

Introduction

Based on advances in sequencing technology, it is well accepted that cancer is generally driven by oncogenic mutations (Park et al. 2013, Kruglyak et al. 2014). Several studies have provided evidence that tumorigenesis strongly correlates with the prevalence of somatic mutations in certain type of cancers (Wood et al. 2007, Vogelstein et al. 2013). The COSMIC (Catalogue of Somatic Mutations in Cancer) database includes hundreds of thousands of human cancer-associated somatic mutations that are classified by tumor type and disease (Forbes et al. 2010). Among the 325,856 mutations registered in the COSMIC database (version 77), more than 90% are substitutions of one or more nucleotides, whereas most of the others are insertions or deletions (indels), indicating that most cancers are caused by simple nucleotide substitutions.

Recently, circulating tumor DNA (ctDNA), which is released into the bloodstream from tumor cells as cell-free DNA fragments, has been proposed as a tumor-specific biomarker candidate (Schwarzenbach et al. 2011, Alix-Panabieres et al. 2012, Freidin et al. 2015, Heitzer et al. 2015, Cheng et al. 2016, De Mattos-Arruda et al. 2016, Sorber et al. 2017). Thus, diagnosing tumors at early stages might be possible by simply detecting tumor-specific somatic mutations in the ctDNA from a patient's blood (Schwarzenbach et al. 2011, Bettegowda et al. 2014). However, cell-free DNAs (cfDNAs) in the blood plasma generally contain extremely small amounts of tumor DNA, which are reportedly

dependent on the tumor burden or cancer stage (Li et al. 2006, Dawson et al. 2013). Therefore, especially in the early stages of cancer, a highly sensitive and specific method would be required to diagnose a tumor by detecting ctDNA (Kinde et al. 2011, Forshew et al. 2012, Narayan et al. 2012, Chan et al. 2013, Newman et al. 2014, Chang-Hao Tsao et al. 2015, Song et al. 2016).

The clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) protein system, an adaptive immune response in prokaryotes, is well known for its specific DNA target recognition and cleavage(Jinek et al. 2012, Kim et al. 2014). The versatility of the CRISPR system has allowed us and other groups to broadly utilize it, not only for genome editing in various organisms (Chang et al. 2013, Cho et al. 2013, Friedland et al. 2013, Wu et al. 2013, Baek et al. 2016), but also to cleave DNA in vitro(Kim et al. 2014). CRISPR endonucleases, such as type II Cas9 or type V Cpf1, specifically induce double-strand breaks in target DNA by recognizing a protospacer-adjacent motif (PAM) downstream or upstream, respectively, of target DNA sequences corresponding to that of a guide RNA (gRNA) (Anders et al. 2014, Nishimasu et al. 2014, Zetsche et al. 2015, Kim et al. 2016, Lee et al. 2016, Yamano et al. 2016).

To use ctDNA for cancer diagnosis, it is necessary to detect, in a highly accurate manner, the small amounts of ctDNAs with missense mutations among the relatively very large amounts of wild-type cfDNAs. By adapting the accurate and specific cleavage ability of CRISPR endonucleases in vitro, it is possible to enrich

tumor-specific versus wild-type alleles by specifically cleaving the wild-type DNAs. Here, I devised a new method employing the CRISPR system, termed CUT (CRISPR - mediated, Ultrasensitive detection of Target DNA) - PCR, that efficiently enriches oncogenic mutant DNAs by eliminating wild-type DNAs prior to PCR amplification. I note that by altering gRNAs corresponding to various wild-type DNAs, one can easily and precisely reduce different background DNA signals in an unbiased manner. After the wild-type DNAs are cleaved by CRISPR endonucleases, the mutant target regions are amplified by PCR and then exquisitely identified by targeted deep sequencing using next generation sequencing (NGS) facilities. Because PCR amplification is performed after background wild-type sequences are reduced, the CUT-PCR process minimizes polymerase-generated errors and maximizes target cleavage specificity.

In this study, I found in silico that CUT-PCR is applicable to more than 80% of the substitution mutations in the COSMIC database. I then validated the specificity and generality of CUT-PCR in vitro by using DNA plasmids to mimic ctDNAs containing human oncogenic mutations. Furthermore, I successfully detected ctDNAs with five recurrent oncogenic mutations among cfDNAs from patients with various stages of colorectal cancer (CRC) with high sensitivity (<0.01%), suggesting that CUT-PCR can be used as a precise early cancer diagnosis method.

Materials and Methods

1. Preparation of CRISPR endonuclease components

Recombinant SpCas9 protein was expressed in *Escherichia coli*. For sgRNA construction, a partial DNA duplex were used as the mother template for *in vitro* RNA transcription. The partial duplex was annealed and then extended by Phusion polymerase (New England Biolabs) and purified with a PCR cleanup kit (DOCTOR PROTEIN, MD008). sgRNA was then generated by *in vitro* transcription at 37 ℃ overnight, after which the template DNA was cleaved with DNA nuclease I (New England Biolabs) at 37 ℃ for 30 minutes. The RNA product was further purified using a PCR cleanup kit. The final RNA concentration was measured with a NanoDrop device (Thermo scientific, NANODROP 2000 spectrophotometer). The purity of the single-stranded RNA was confirmed with 8% urea-PAGE (>95% purity). The wild-type FnCpf1 protein and crRNA were purified in the same manner.

2. Selection of oncogene candidates

I found candidate recurrent missense mutations in the COSMIC database with *in silico* based research. The most frequent mutations that affected PAM sites in potential guide RNA target sequences, such that they cannot be targeted by wild-type specific sgRNAs, were selected.

3. Construction of plasmids containing wild-type and mutant sequences

PCR amplicons of relevant wild-type proto-oncogene cDNAs were subcloned into the commercial T-blunt cloning vector (T-Blunt^{TM} PCR cloning Kit, SolGent) using the manufacturer's protocol. Single base-pair substituted mutations (KRAS: c.35G>A, c.35G>T, c.34G>T, c.35G>C, c.34G>C, GNAQ: c.626A>T) were constructed by site-directed mutagenesis using appropriate primer sets. For EGFR vector construction, cDNAs of wild-type EGFR and its deletion mutant (c.2235-2249del) were subcloned into the T-blunt vector. All mutations in the cloned target sequences were confirmed by Sanger sequencing.

4. *In vitro* DNA cleavage assay to test PAM recognition

To test plasmid cleavage by CRISPR nucleases, target plasmids were first linearized with the restriction enzyme NcoI (New England Biolabs) for 37 ℃, 1 hour (10ul reaction in NEB buffer 3.1). The linearized product was further cleaved by treatment with a wild-type sequence-specific CRISPR nuclease (Cas9 100 ng, sgRNA 70 ng, 10 ul reaction in NEB buffer 3.1 at 37 ℃, 1 hour).

5. Cell-free DNA extraction from plasma of CRC patients and healthy donors

Peripheral blood samples from patients were obtained from the Pusan National University Hospital (Busan, Korea). This study was

reviewed and approved by the Institutional Review Board (IRB) of PNUH(H-1412-011-024) and UNIST(UNISTIRB-13-002-A). To get cell-free DNA from CRC patients and healthy volunteers, plasma was obtained from blood sample by using Ficoll-PaqueTM PLUS (GE Healthcare) and Cell-free DNA was purified from 1 mL of the plasma with a QIAamp® Circulating Nucleic Acid kit (Qiagen) according to the manufacturer's protocol.

6. Enrichment of mutant alleles through CUT-PCR

To test the ability of CUT-PCR to enrich mutant alleles, DNA plasmids containing wild-type and mutant sequences were mixed in various ratios and subjected to CRISPR cleavage in vitro. The plasmid mixture was treated with a wild-type specific CRISPR nuclease (11 ng total plasmid DNA, 100 ng Cas9, 70 ng sgRNA, 10 ul reaction in NEB buffer 3.1 at 37 ℃ for 1hour) to cleave wild-type DNA. After proteinase (Qiagen) treatment, samples were purified using a PCR cleanup kit (DOCTOR PROTEIN, MD008) and each sample was amplified by PCR using targeted primer sets. To quantify target-specific cleavage, the fold increase in the target sequence was compared to that of an internal control product, which was amplified with internal control primers. For CUT-PCR using cfDNA from plasma samples, CRISPR cleavage was conducted using conditions similar to those described for plasmid DNA (1ng total cfDNA, 10 ng Cas9, 7 ng sgRNA, 10 ul reaction in NEB buffer 3.1 at 37 ℃ for 1hour), except that multiple rounds of the cleavage reaction were required to enrich the extremely small amounts of

cfDNA.

7. Real-time quantitative PCR

Real-time quantitative PCR (qRT-PCR) analysis was performed in 96-well format using the CFX 96 Touch Real-Time PCR Detection System (Bio-Rad) and iQ supermix (Bio-Rad). 1 ng of mixed plasmid were used in a 20 μl reaction. Probes that targeted KRAS wild-type and 35G>T mutation were labeled with FAM and Cy5. Reaction conditions were as follows: 3 min at 95 ℃, 45 cycles of PCR (10 sec at 95 ℃, 30 sec at 58 ℃, 15 sec at 72 ℃), Data was collected at 72 ℃ in each cycle. The qRT-PCR analysis was technically repeated three times.

8. Targeted deep sequencing and data analysis

CUT-PCR enriched plasmid and cfDNAs were further amplified with adaptor primers using Phusion polymerase (New England Biolabs). The resulting PCR amplicons were subjected to paired-end sequencing with the Illumina MiSeq system. After MiSeq, paired-end reads were joined by the Fastq-join, a part of ea-utils program (https://code.google.com/archive/p/ea-utils/) using default values. Paired-end reads were then analyzed by comparing wild-type and mutant sequences using Cas-Analyzer.

Table 1. Primers used in this study.

Primer	Sequence (5' to 3')
*KRAS*_ forward primer	AACAATTTCACACAGGAAACAGCT
*KRAS*_ reverse primer	TACGATTCACTATAGGGCGAATTG
KRAS internal control forward primer	TTTCCGTGTCGCCCTTATTCC
KRAS internal control reverse primer	TGTCATGCCATCCGTAAGATGC
*KRAS*_c.35G>A mutation forward primer	ACTTGTGGTAGTTGGAGCTGATGGCGTAGGCAAGAGTGCCT
*KRAS*_c.35G>A mutation reverse primer	AGGCACTCTTGCCTACGCCATCAGCTCCAACTACCACAAGT
*KRAS*_c.35G>T mutation forward primer	ACTTGTGGTAGTTGGAGCTGTTGGCGTAGGCAAGAGTGCCT
*KRAS*_c.35G>T mutation reverse primer	AGGCACTCTTGCCTACGCCAACAGCTCCAACTACCACAAGT
*KRAS*_c.34G>T mutation forward primer	AACTTGTGGTAGTTGGAGCTTGTGGCGTAGGCAAGAGTGCC
*KRAS*_c.34G>T mutation reverse primer	GGCACTCTTGCCTACGCCACAAGCTCCAACTACCACAAGTT
*KRAS*_c.35G>C mutation forward primer	ACTTGTGGTAGTTGGAGCTGCTGGCGTAGGCAAGAGTGCCT
*KRAS*_c.35G>C mutation reverse primer	AGGCACTCTTGCCTACGCCAGCAGCTCCAACTACCACAAGT
*KRAS*_c.34G>C mutation forward primer	AACTTGTGGTAGTTGGAGCTCGTGGCGTAGGCAAGAGTGCC
*KRAS*_c.34G>C mutation reverse primer	GGCACTCTTGCCTACGCCACGAGCTCCAACTACCACAAGTT
KRAS adaptor forward primer (for cell-free DNA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTGTGACATGTTCTAATATAGTCACATTTT
KRAS adaptor reverse primer (for cell-free DNA)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTCTGAATTAGCTGTATCGTCAAGGCACTCTT
KRAS adaptor forward primer (for plasmid DNA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTGTATTAACCTTATGTGTGACATGTTCT
KRAS adaptor reverse primer (for plasmid DNA)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGAGAAACCTTTATCTGTATCAAAGAATG

(continued)

Primer	Sequence (5' to 3')
Internal control *CCR5* forward primer	GAGGGCAACTAAATACATTCTAGGAC
Internal control *CCR5* reverse primer	GCCTTTTGCAGTTTATCAGGATGAGG
KRAS forward primer	TGTGTGACATGTTCTAATATAGTCACAT
KRAS reverse primer	TTCTGAATTAGCTGTATCGTCAAGGCAC
KRAS wild-type sgRNA forward primer	GAAATTAATACGACTCACTATAGGAAACTTGTGGTAGTTGGAGCGTTTTAGAGCTAGAAATAGCAAG
KRAS wild-type sgRNA reverse primer	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC
GNAQ_ forward primer	CATGATGTGTTACCCAGAATGTTTTAAC
GNAQ_ reverse primer	CCACACCCTACTTTCTATCATTTACTTG
GNAQ_internal control forward primer	TTCCTGTTTTTGCTCACCCAG
GNAQ_ internal control reverse primer	TCTGAGAATAGTGTATGCGGCG
GNAQ_ c626A>T mutation forward primer	GCAGAATGGTCGATGTAGGGGGCCTAAGGTCAGAGAGAAGAAAAT
GNAQ_c626A>T mutation reverse primer	ATTTTCTTCTCTCTGACCTTAGGCCCCCTACATCGACCATTCTGC
CTNNB1_ forward primer	AAACAAGCCACCAGCAGGAATCT
CTNNB1_ reverse primer	TTCAACACTCACTATCCACAGTTCAGC
CTNNB1_c.133T>C mutation forward primer	TGGTGCCACTACCACAGCTCCTCCTCTGAGTGGTAAAGGCAATCCT
CTNNB1_ c.133T>C mutation reverse primer	AGGATTGCCTTTACCACTCAGAGGAGGAGCTGTGGTAGTGGCACCA
EGFR_ forward primer	CCACACGGACTTTATAACAGGCTTTACA
EGFR_ reverse primer	CATGCTCCAGGCTCACCAAGAGCACTC

(continued)

Primer	Sequence (5' to 3')
GNAQ adaptor forward primer (for plasmid DNA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTAAGTTTGTAAGTAGTGCTATATTTA
GNAQ adaptor reverse primer (for plasmid DNA)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCATTGTCTGACTCCACGAGAACTTGA
CTNNB1 adaptor forward primer (for plasmid DNA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCCATGGAACCAGACAGAAAAGCGGCT
CTNNB1 adaptor reverse primer (for plasmid DNA)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCTACTTGTTCTTGAGTGAAGGACTG
EGFR adaptor forward primer (for plasmid DNA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCAGCATGTGGCACCATCTCACAATTGCCAGT
EGFR adaptor reverse primer (for plasmid DNA)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATAAGTGAACATTTAGGATGTGGAGATGAGCA
GNAQ wild-type crRNA forward primer	GAAATTAATACGACTCACTATAGGG
GNAQ wild-type crRNA reverse primer	TGCAGAATGGTCGATGTAGGGGGCATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTA ATTTC'
CTNNB1 wild-type crRNA forward primer	GAAATTAATACGACTCACTATAGGG
CTNNB1 wild-type crRNA reverse primer	TCAGGATTGCCTTTACCACTCAGAATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAA TTTC'
EGFR wild-type sgRNA forward primer	GAAATTAATACGACTCACTATAGGCTTAATTCCTTGATAGCGACGTTTTAGAGCTAGAAATAGC AAG
EGFR wild-type sgRNA reverse primer	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC TATTTCTAGCTCTAAAAC

Results

1. *In silico* investigation of applicable targets for CUT-PCR in the COSMIC database.

By using the specific recognition property of CRISPR endonucleases for DNA PAM sequences, it is possible to cleave PAM-containing wild-type DNA sequences selectively, reducing background DNA signals and enriching cancer-specific mutant DNA signals (Figure 1A). The representative type II CRISPR endonuclease Cas9, derived from *Streptococcus pyogenes* (SpCas9), and type V Cpf1, from *Francisella novicida* (FnCpf1), respectively recognize the PAM sequences 5'-NGG-3', located downstream of the target DNA (Anders et al. 2014), and 5'-TTN-3', located upstream of the target DNA (Zetsche et al. 2015). Therefore, if oncogenic mutant sequences are generated by single-base substitutions such as these (NGG>NGH or NGG>NHG, H is A, C, or T; TTN>TVN or TTN>VTN, V is A, C, or G) in wild-type DNA sequences, wild-type DNAs can be selectively and precisely eliminated by SpCas9 or FnCpf1, respectively (Jinek et al. 2012, Zetsche et al. 2015). After wild-type DNA cleavage, pooled DNAs can be identified directly by PCR amplification followed by targeted deep sequencing (Park et al. 2017).

I inspected all possible target sites registered in the COSMIC database in silico to determine whether various orthologonal Cas9 or Cpf1 proteins would be applicable for the specific destruction of the corresponding wild-type sequences. As shown in Figure 1B,

among 325,856 mutations registered in the COSMIC database (version 77), 90.3% are single- or multiple-nucleotide substitutions. Insertions and deletions account for 2.9% and 6.6% of the entries, respectively, and unknown patterns represent 0.2%. For each indel, I searched for an adjacent PAM sequence to design sgRNA because CRISPR endonucleases barely cleave mutant DNA that contains indels in the gRNA target region (Kim et al. 2014, Zetsche et al. 2015). In the case of the substitutions, however, CRISPR endonucleases can typically cleave mutant DNAs as well as wild-type DNAs in vitro because of their ability to recognize sites that vary by one or a few nucleotides from the gRNA sequence (off-target effects) (Cho et al. 2014). But PAM recognition by CRISPR endonucleases is much stricter; they rarely cleave target DNA that lacks a PAM sequence even if the target DNA is exactly complementary to the gRNA. Thus, I determined if each substitution mutation might have destroyed a PAM sequence in the corresponding wild-type DNA, which would mean that the wild-type DNA would be cleaved much more readily than the oncogenic DNA that lacked a proper PAM. Further analyses resultantly showed that 98.9% of the indels (Table 2) and 80.5% of the substitutions (Table 3) represent DNA targets that can be selectively cleaved by the various orthologonal Cas9 or Cpf1 proteins reported to date, suggesting that CUT-PCR method would be useful for detecting about 80% of the oncogenic mutations in the COSMIC database (Figure 1C).

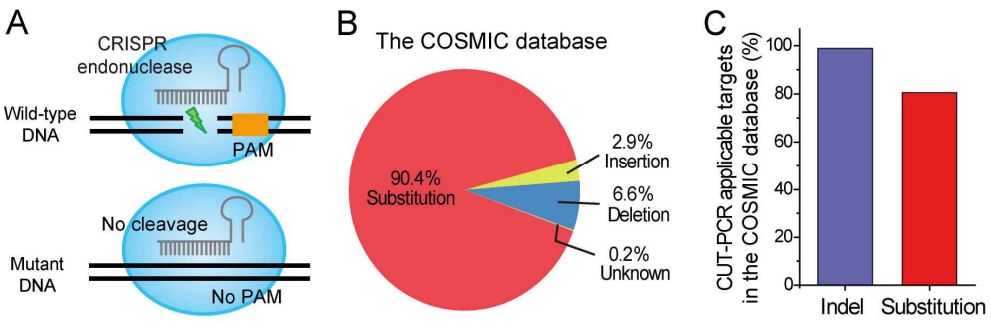


Figure 1. CUT-PCR method and its applicable targets in the COSMIC database. (A) Schematic of the CUT-PCR enrichment process. To cleave wild-type DNA specifically, single-guide RNAs were designed to target PAM sites that are destroyed by oncogenic mutations. Such mutant alleles are not recognized by the CRISPR endonucleases and largely avoid cleavage. After cleavage of wild-type DNA, the DNA in the pooled solution was amplified with PCR. (B) The classification of human cancer-associated somatic mutations registered in the COSMIC database. (C) The ratio of CUT-PCR applicable targets among the mutations of indel (blue bar) and substitution (red bar) registered in the COSMIC database.

Table 2. Various CRISPR applicable target ratios among indel mutations in the COSMIC database for CUT-PCR.

CRISPR Type	Related organism	PAM sequence (5' to 3')	Frequency	Ratio for substitution mutation (%)
Type II Cas9 endonuclease	*S.pyogenes* VRER	GCG	14597	47.33753
	S.pyogenes	GG	28769	93.2968
	S.mutans	GG	29385	95.29446
	*S.thermophilus*_1	GGNG	11494	37.27461
	*S.thermophilus*_2	AGAAW	9403	30.49358
	C.jejuni	ACA	19202	62.27137
	N.meningitidis	GATT	7983	25.88857
	P.multocida	GNNNCNNA	7932	25.72318
	F.novicida	GG	28912	93.76054
	P.lavamentivorans	CAT	18724	60.72123
	S.pasteurianus	GTGA	10149	32.91283
	N.cinerea	GTA	12574	40.77701
Type V Cpf1 endonuclease	*S.aureus*	GRRT	8942	28.99857
	C.lari	GGG	18423	59.7451
	C.diphtheriae	GG	28769	93.2968
	F.novicida U 112	TT	25611	83.05552
	Acidaminococcus sp. BV2L6	TTT	18485	59.94617
	Lachnospiraceae bacterium MA2020	TTT	18485	59.94617
	Candidatus methanoplasma termitum	TTTTA	5532	17.94007
	Moraxella bovoculi 237	TT	25611	83.05552
	Lachnospiraceae bacterium ND2006	TTT	18485	59.94617
	Porphyromonas crevioricanis	TTTA	18485	59.94617
	Prevotella disiens	TTTTM	5532	17.94007

Table 3. Various CRISPR applicable target ratios among substitution mutations in the COSMIC database for CUT-PCR.

CRISPR Type	Related organism	PAM sequence (5' to 3')	Frequency	Ratio for substitution mutation (%)
Type II Cas9 endonuclease	*S.pyogenes* VRER	GCG	22348	6.858244
	S.pyogenes	GG	78403	24.06063
	S.mutans	GG	78403	24.06063
	*S.thermophilus*_1	GGNG	7271	2.231354
	*S.thermophilus*_2	AGAAW	2932	0.899784
	C.jejuni	ACA	90317	27.71684
	N.meningitidis	GATT	1642	0.503904
	P.multocida	GNNNCNNA	39550	12.13726
	F.novicida	GG	78403	24.06063
	P.lavamentivorans	CAT	11617	3.565072
	S.pasteurianus	GTGA	28381	8.709675
	N.cinerea	GTA	6067	1.861865
Type V Cpf1 endonuclease	*S.aureus*	GRRT	10669	3.274146
	C.lari	GGG	12410	3.808431
	C.diphtheriae	GG	78403	24.06063
	F.novicida U 112	TT	10354	3.177477
	Acidaminococcus sp. BV2L6	TTT	4200	1.288913
	Lachnospiraceae bacterium MA2020	TTT	4200	1.288913
	Candidatus methanoplasma termitum	TTTTA	248	0.076107
	Moraxella bovoculi 237	TT	10354	3.177477
	Lachnospiraceae bacterium ND2006	TTT	4200	1.288913
	Porphyromonas crevioricanis	TTTA	4200	1.288913
	Prevotella disiens	TTTTM	734	0.225253

2. Validation of PAM recognition specificity of CRISPR endonucleases using plasmids.

To validate that CRISPR endonucleases could selectively cleave target DNA as specified in the CUT-PCR protocol, I performed *in vitro* cleavage assays with T-vector cloned sequences containing various missense mutations. I expected that CRISPR endonucleases with gRNAs specific to the wild-type sequence would specifically deplete the wild-type DNA. I chose five recurrent cancer-associated mutations in the *KRAS* gene (*KRAS* c.35G>A, c.35G>T, c.34G>T, c.35G>C, c.34G>C) for testing type II SpCas9 and one in the *GNAQ* gene (GNAQ c.626A>T) for testing type V FnCpf1. Both oncogenes are well known for their tumorigenicity (Xu et al. 2014, Park et al. 2015).

For testing the SpCas9, I constructed one plasmid containing the wild-type *KRAS* sequence and five plasmids containing patient-mimic sequences in which the PAM sequence was changed, as shown in Figure 2A. I validated that each plasmid can be linearized with NcoI restriction enzyme and then treated each linearized plasmid in vitro with the SpCas9 complex containing a single-guide RNA (sgRNA) specific to the wild-type sequence. As a result, it showed that the SpCas9 complex selectively cleaved wild-type DNA resulting in shorter DNA fragments but generally did not cleave the other mutant sequences that lacked a functional PAM (Figure 2B). I note that one mutant plasmid containing *KRAS* (c.35G>A), which has a 5' -NGA-3' PAM sequence, is

marginally targeted by SpCas9 specific for the wild-type sequence, a result already reported in a previous study (Zhang et al. 2014).

For testing the FnCpf1 nuclease, I constructed two different plasmids, one containing the wild-type *GNAQ* sequence and the other the patient-mimic mutated sequence, as shown in Figure 2C. Both plasmids can be also linearized with NcoI. In this case, I designed one CRISPR RNA (crRNA) specific to the wild-type sequence and treated each plasmid in vitro with the FnCpf1 complex. Results showed that the FnCpf1 complex selectively cleaved wild-type plasmid DNA but not the mutant sequence that lacked a functional PAM (Figure 2D), suggesting that FnCpf1 would also be useful for mutant sequence enrichment in the CUT-PCR process.

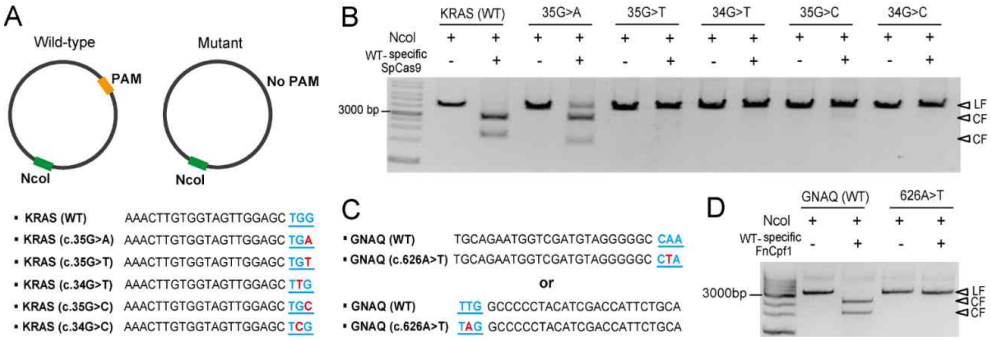


Figure 2. *In vitro* cleavage assay with plasmids containing sequences with PAM mutations. (A) Top: Schematic of the plasmids containing sequences with wild-type and oncogenic mutations. Plasmids can be linearized by the restriction enzyme NcoI. Bottom: The sequences of wild-type and recurrent *KRAS* mutations in the COSMIC database. The PAM sequence (5'-TGG-3') for SpCas9 is underlined in blue. Missense mutations are shown in red. (B) *In vitro* cleavage assay using SpCas9 with linearized plasmids containing wild-type and mutant KRAS sequences. LF: linearized fragment, CF: cleaved fragment. (C) The sequences of wild-type and recurrent *GNAQ* mutation in the COSMIC database. The PAM sequence (5'-TTG-3') for FnCpf1 is underlined in blue. (D) *In vitro* cleavage assay using FnCpf1 with linearized plasmids containing wild-type and mutant *GNAQ* sequences.

3. CUT-PCR based enrichment of sequences with missense mutations.

To investigate whether CUT-PCR could be used to detect rare oncogene-specific mutations, I next prepared mixtures in which the plasmid containing a mutant sequence was serially diluted with the plasmid containing the wild-type sequence. I then treated each plasmid mixture *in vitro* with a CRISPR endonuclease and a gRNA specific to the wild-type sequence, after which I amplified the target region using PCR (Figure 3A). I expected that plasmids containing the wild-type sequence would be selectively cleaved by CRISPR complexes, resulting in relatively less amplification, whereas plasmids containing the mutant sequences would not be cleaved and would therefore be amplified more. When a mixture of plasmids containing either wild-type or mutant *KRAS* (c.35G>T) sequences was treated with wild-type specific SpCas9 complexes in vitro, the total amount of PCR amplicons gradually decreased as the abundance of the *KRAS* mutant plasmid decreased (Figure 3B, lanes 1-5), in contrast to untreated samples (Figure 3B, lanes 6-10). This result indicates that most wild-type sequences were eliminated by SpCas9, suggesting that mutant sequences were enriched relative to the wild-type sequences in the mixture of PCR amplicons. As a quantitative control for PCR amplification in each reaction, I added pairs of internal control primers to each mixture and determined the PCR outcomes relative to these control PCR products.

To examine the sensitivity of CUT-PCR method, I conducted targeted deep-sequencing for each mixture with various ratios of mutant plasmids using Illumina MiSeq. Every sample was read at a sequencing depth of at least 10000x. I sought to compare CUT-PCR based deep sequencing against the conventional targeted deep sequencing. For *KRAS* (c.35G>T) sample, mutant plasmids were originally mixed with wild-type plasmids at a ratio of from 100% to 0.01%. Then, DNA target sites were amplified by PCR after CUT-PCR treated or not. As a result, conventional deep sequencing for missense mutations was limited to 0.1% as in previous study (Cho et al. 2014). However, mutant DNA fragments were entirely enriched after SpCas9-based CUT-PCR treatments (Figure 3C). For the mixture that mutant plasmids were mixed at a ratio of 0.01%, CUT-PCR treated samples showed a 6-fold increase in the mutated DNA fragment frequency, relative to untreated samples. For the fold enrichment calculation, I used the value from the CUT-PCR untreated sample as the background frequency. In addition, the mutated DNA fragment frequencies were more increased (Figure 3D) after multiple rounds of CUT-PCR, resulting in a greater fold increase (Figure 3E). These results indicate that the sensitivity of CUT-PCR based deep sequencing is more than 0.01%. For the additional comparison with quantitative real time PCR (qPCR), it is hard to detect missense mutations among the mixtures at a ratio of 1% of mutant plasmids (Figure 4).

I repeated the CUT-PCR procedure with other *KRAS* mutant plasmid mixtures as described above and obtained the same results

as in the results for *KRAS* (c.35G>T) shown in Figure 3 and 5. For the mixture that mutant plasmids were originally mixed at a ratio of 0.1%, mutant DNA fragments were significantly enriched by SpCas9-based CUT-PCR relative to untreated samples (Figure 6A). And the values of a fold increase in the mutated DNA fragment frequency were from 29.6 to 76.3 (Figure 6B). I note that in the case of the *KRAS* (c.35G>A) mutant sequence, some of the plasmids were cleaved by wild-type specific SpCas9 as shown in Figure 2B, but the relative amount of mutant DNA fragments was strongly increased after CUT-PCR, which might indicate that wild-type DNA fragments were preferentially eliminated.

I further tested whether CUT-PCR enrich mutant DNA for different target sites and different CRISPR type. For the *GNAQ* (c.626A>T) mutant and wild-type plasmid mixture, I verified that FnCpf1 would cleave the wild-type DNA fragment selectively and sufficiently (Figure 7). In line with SpCas9-based CUT-PCR, I determined that a mixture treated with FnCpf1-based CUT-PCR showed a 27-fold increase in the mutant fragment frequency as compared with untreated samples when *GNAQ* mutant and wild-type plasmids were originally mixed at a ratio of 0.1% (Figure 6C and 6D). To test whether CUT-PCR could be used more generally, I applied the process to other oncogenes. I used FnCpf1-based CUT-PCR with *CTNNB1* containing a substitution mutation (Figure 8) and SpCas9-based CUT-PCR with *EGFR* containing a deletion (Figure 9). I also found that one cycle of CUT-PCR efficiently enriched mutant DNA fragments as in the above results.

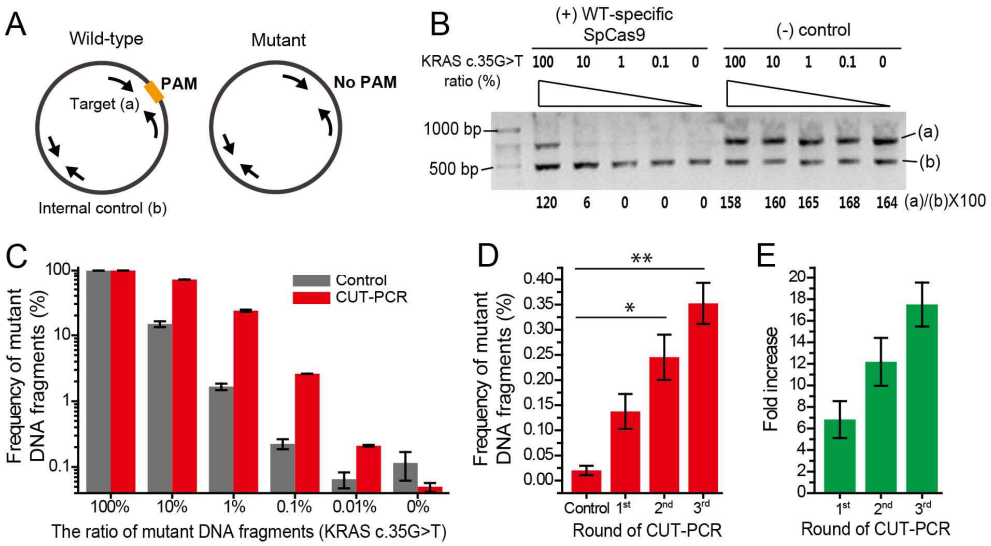
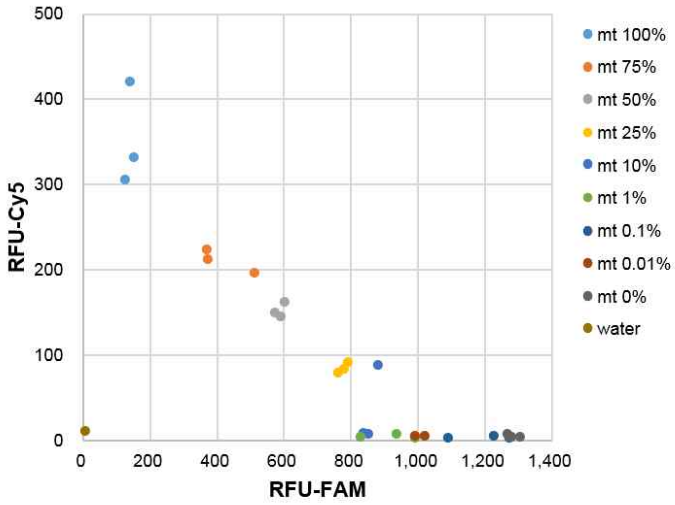


Figure 3. CUT-PCR based enrichment of plasmid-borne sequences containing *KRAS* (c.35G>T) mutation. (A) Schematic of two sets of primers for target (929nt) and internal control region (334nt). After treatment of wild-type specific CRISPR endonucleases, each plasmid mixture containing sequences with wild-type and oncogenic mutation was amplified using PCR. (B) CUT-PCR experiment for various ratios of plasmid mixtures containing either wild-type or mutant *KRAS* (c.35G>T) sequence. The amount of the amplified *KRAS* target region relative to the internal control PCR product in each lane was calculated. (C) For the recurrent *KRAS* (c.35G>T) mutation, targeted deep sequencing after CUT-PCR was treated (red bars) or not (gray bars) were conducted for the plasmid mixtures in which mutant plasmids were originally mixed with wild-type plasmids at a ratio of from 100% to 0.01%. For the mixture at a ratio of 0.01%, frequencies of mutant DNA fragments (D) were measured by targeted deep sequencing and the values of

fold increase (E) were calculated after multiple round of CUT-PCR treatments. (Error bars mean s.e.m.; n=2 for C and 3 for D; * $P < 0.05$; ** $P < 0.01$)



KRAS Q-PCR forward primer	5'-AGGCCTGCTGAAAATGACTG-3'
KRAS Q-PCR reverse primer	5'-GCTGTATCGTCAAGGCACTCT-3'
KRAS wild-type probe	5'-GTTGGAGCTGGTGGCGTAG-3'
KRAS mutant (35G>T) probe	5'-GTTGGAGCTGTTGGCGTAG-3'

Figure 4. Real-time quantitative PCR using selective sensitivity probes. *KRAS* mutant plasmids (35G>T) were mixed with wild-type plasmids at a ratio from 100% to 0%. 1 ng of mixed plasmid were used in a 20 μl PCR reaction. The qRT-PCR analysis was performed in 96-well format using the CFX 96 Touch Real-Time PCR Detection System (Bio-Rad) and iQ supermix (Bio-Rad). Wild-type probe contained 5' FAM and 3' BHQ1, mutant probe contained 5' cy5 and 3' BHQ3. Primers and probes used for Real-time quantitative PCR were shown at the bottom as a table. Reaction conditions were as follows: 3 min at 95℃, 45 cycles of PCR (10 sec at 95℃, 30 sec at 58℃, 15 sec at 72℃), Data was collected at 72℃ in each cycle. The qRT-PCR analysis was technically repeated three times.

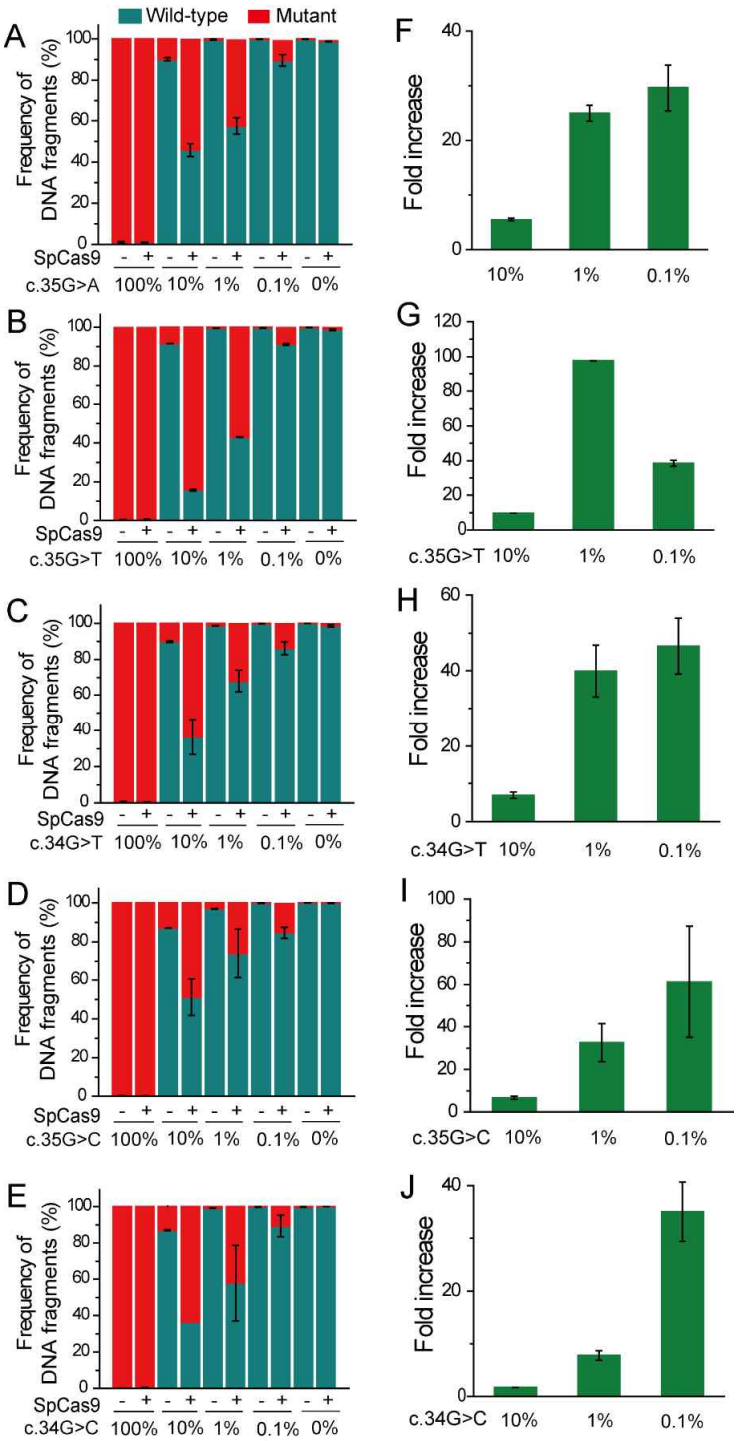

A
Wild-type
Mutant
Frequency of DNA fragments (%)
SpCas9
c.35G>A 100% 10% 1% 0.1% 0%
F
Fold increase
10% 1% 0.1%
B
Frequency of DNA fragments (%)
SpCas9
c.35G>T 100% 10% 1% 0.1% 0%
G
Fold increase
c.35G>T 10% 1% 0.1%
C
Frequency of DNA fragments (%)
SpCas9
c.34G>T 100% 10% 1% 0.1% 0%
H
Fold increase
c.34G>T 10% 1% 0.1%
D
Frequency of DNA fragments (%)
SpCas9
c.35G>C 100% 10% 1% 0.1% 0%
I
Fold increase
c.35G>C 10% 1% 0.1%
E
Frequency of DNA fragments (%)
SpCas9
c.34G>C 100% 10% 1% 0.1% 0%
J
Fold increase
c.34G>C 10% 1% 0.1%

Figure 5. CUT-PCR experiments for various ratios of plasmid mixtures containing wild-type or different mutant *KRAS* sequences. For the five recurrent *KRAS* mutations (A) c.35G>A, (B) c.35G>T, (C) c.34G>T, (D) c.35G>C, (E) c.34G>C, the frequencies of wild-type (blue bars) and mutant (red bars) fragments were measured using targeted deep sequencing. Fold increases after CUT-PCR in *KRAS* mutant DNA ((F) c.35G>A, (G) c.35G>T, (H) c.34G>T, (I) c.35G>C, (J) c.34G>C) frequency were calculated from the deep sequencing data of (A) – (E). Each background value was obtained by averaging the mutant DNA frequency of CUT-PCR untreated sample.

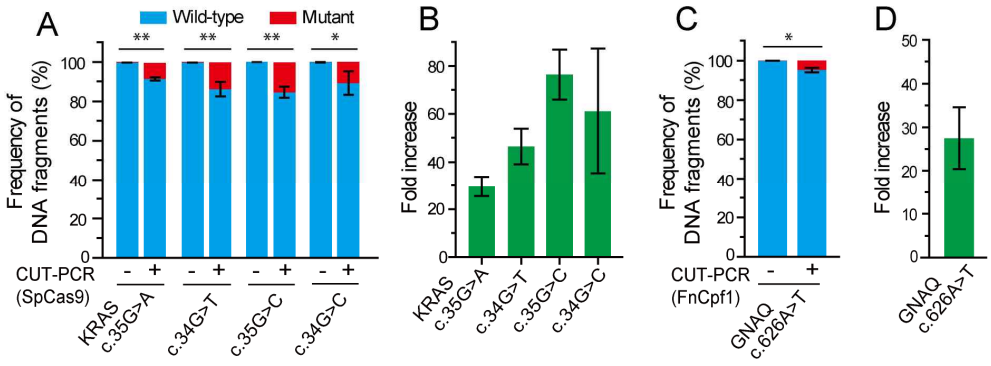


Figure 6. CUT-PCR based enrichment of plasmid-borne sequences containing recurrent cancer mutations. (A) For the four recurrent *KRAS* mutations (c.35G>A, c.34G>T, c.35G>C, c.34G>C), the frequencies of wild-type (blue bars) and mutant (red bars) fragments were measured using targeted deep sequencing. In all cases, *KRAS* mutant plasmids were originally mixed with wild-type plasmids at a ratio of 0.1%. (B) Fold increase after CUT-PCR in each *KRAS* mutant DNA frequency was calculated from the data of (A). (C) DNA frequencies of wild-type and *GNAQ* mutant (c.626A>T) fragments measured using targeted deep sequencing after FnCpf1 mediated CUT-PCR. *GNAQ* mutant plasmids were originally mixed with wild-type plasmids at a ratio of 0.1%. (D) Fold increase in the mutant DNA fragment frequency for the recurrent *GNAQ* mutation calculated from (C). (Error bars mean s.e.m.; n= 3; * $P < 0.05$; ** $P < 0.01$)

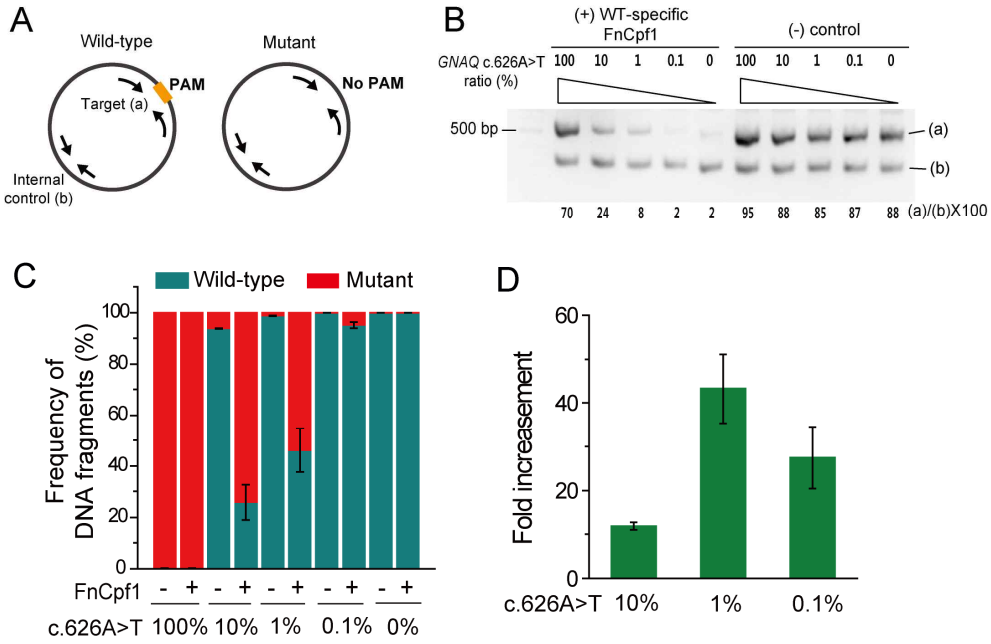


Figure 7. CUT-PCR experiment for various ratios of plasmid mixtures containing either wild-type or mutant *GNAQ* (c.626A>T) sequence. (A) Schematic of two sets of primers for target (412bp) and internal control region (237bp). (B) After FnCpf1 nucleases specific for wild-type *GNAQ* had been treated in vitro (lanes 1 to 5) or not (lanes 6 to 10), PCR amplicons from the mixture of plasmids were separated on a 2% agarose gel. The amount of the amplified *GNAQ* target region relative to the internal control PCR product in each lane was calculated. The *GNAQ* mutant (c.626A>T) plasmid was serially diluted (up to 0%) with wild-type plasmid DNA before CUT-PCR. (C) For the various ratios of plasmid mixtures, the frequencies of wild-type (blue bars) and mutant (red bars) fragments were measured using targeted deep sequencing. (D) Fold increases after CUT-PCR in *GNAQ* mutant (c.626A>T)

frequency were calculated from the deep sequencing data of (C). Each background value was obtained by averaging the mutant DNA frequency of CUT-PCR untreated sample.

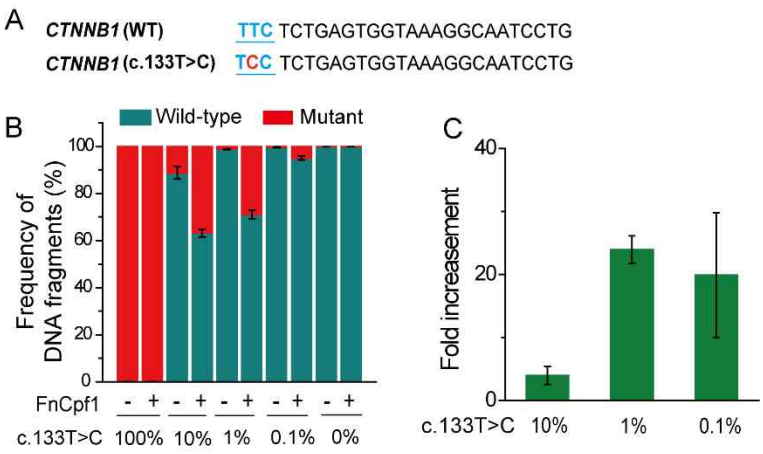


Figure 8. CUT-PCR experiment for various ratios of plasmid mixtures containing either wild-type or mutant *CTNNB1* (c.133T>C) sequence. (A) The sequences of wild-type and recurrent *CTNNB1* mutation in the COSMIC database. The PAM sequence (5'-TTC-3') for FnCpf1 is underlined in blue. Missense mutations are shown in red. (B) For the various ratios of plasmid mixtures, the frequencies of wild-type (blue bars) and mutant (red bars) fragments were measured using targeted deep sequencing. The *CTNNB1* mutant (c.133T>C) was serially diluted (up to 0%) with wild-type plasmid DNA before CUT-PCR. (C) Fold increases after CUT-PCR in *CTNNB1* mutant (c.133T>C) frequency were calculated from the deep sequencing data of (B). Each background value was obtained by averaging the mutant DNA frequency of CUT-PCR untreated sample.

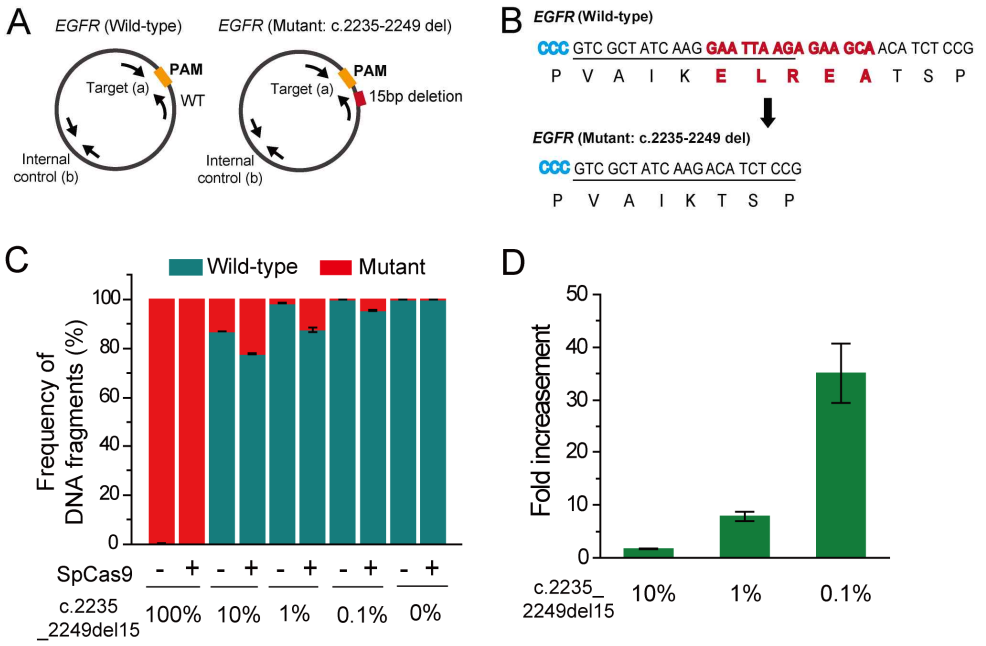


Figure 9. CUT-PCR experiment for various ratios of plasmid mixtures containing either wild-type or deleted *EGFR* (c.2235-2249del) sequence. (A) Schematic of the plasmids containing sequences with wild-type and oncogenic mutations. For mutant plasmid, 15 bases were deleted compared with wild-type. Arrows indicate the two sets of primers for target (929nt) and internal control region (334nt). (B) The sequences of wild-type and *EGFR* (c.2235-2249del) mutation in the COSMIC database. The PAM sequence (5'-CCC-3') for SpCas9 is colored in blue and deleted sequences are shown in red. (C) For the various ratios of plasmid mixtures, the frequencies of wild-type (blue bars) and mutant (red bars) fragments were measured using targeted deep sequencing. The *EGFR* mutant (c.2235-2249del) was serially diluted (up to 0%) with wild-type plasmid DNA before CUT-PCR. (D) Fold

increases after CUT-PCR in *EGFR* mutant (c.2235-2249del) frequency were calculated from the deep sequencing data of (C). Each background value was obtained by averaging the mutant DNA frequency of CUT-PCR untreated sample.

4. Enrichment and detection of ctDNA in the plasma of CRC patients.

I ultimately applied the CUT-PCR technique to detect cell-free ctDNA extracted from the blood plasma of eight CRC patients at various stages of disease. As a control, I used plasma from four healthy donors. *KRAS* mutations are frequently found in colon cancer. To enrich mutant *KRAS* ctDNAs, I prepared sgRNA for SpCas9 specific to the wild-type *KRAS* sequence as described above. As shown in Figure 3 and 4, *KRAS* sequences containing five different oncogenic mutations (KRAS c.35G>A, c.35G>T, c.34G>T, c.35G>C, and c.34G>C), can be enriched using one common sgRNA because of PAM sequence substitution. Because total amounts of cfDNAs in plasma are low and mutant ctDNA fragments are present at very low abundance, especially at early stages of disease, I conducted multiple rounds of CUT-PCR. After each round of CUT-PCR, I measured mutant and wild-type *KRAS* allele frequencies (AFs) by targeted deep-sequencing (Figure 10A). After the third round of CUT-PCR, I measured the enriched mutant AF and calculated its fold increase relative to the wild-type AF (Figure 10B).

To determine the baseline of background mutations detected in multiple rounds of CUT-PCR, I analyzed four healthy donor specimens as a control. For the control samples, no significant increase in mutant AFs was observed after multiple rounds of CUT-PCR (Figure 10A, blue region). In parallel to the healthy

controls, I performed three rounds of CUT-PCR with eight cfDNA samples from CRC patients. All experiments were repeated three times, after which I calculated mean values and standard deviations. For the (*KRAS* c.35G>A) mutation, as shown in the pink regions in the upper panel of Figure 10A, I found that mutant AFs from patients 2, 3, 4, 5 and 7 increased considerably compared with the average value from the healthy controls. During the third round of CUT-PCR, the mutant AFs from patients' samples increased from 164 to 640 fold as compared to the control that lacked SpCas9 treatment (Figure 10B, pink region). I calculated the fold increase of mutant AFs in the same way for the other mutations and summarized the results in Table 4. In conclusion, our CUT-PCR data from cfDNAs provide as much information as the pyrosequencing data from tissue samples. Furthermore, I obtained more information about mutation patterns even in the patient with Stage I cancer.

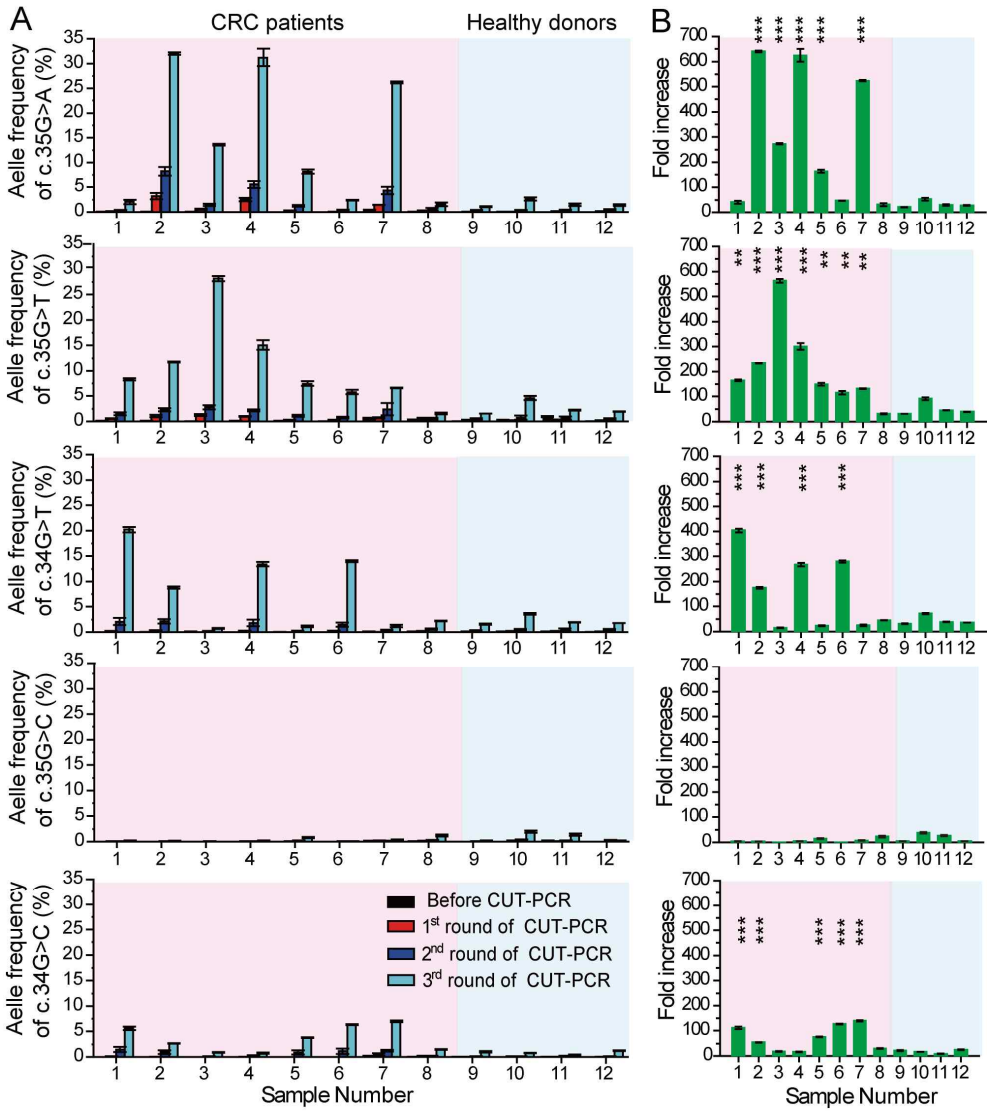


Figure 10. CUT-PCR based enrichment of sequences containing recurrent *KRAS* mutations in cfDNAs from CRC patients and healthy donors. (A) The allele frequencies (AFs) of recurrent *KRAS* mutation candidates (c.35G>A, c.35G>T, c.34G>T, c.35G>C, c.34G>C) were analyzed from cfDNAs in plasma of eight CRC patients (pink boxes) and four healthy donors (blue boxes). After the multiple rounds of CUT-PCR treatment using the wild-type

KRAS specific SpCas9 nucleases, each AF of recurrent *KRAS* mutation sequence was measured using targeted deep sequencing for all samples, respectively. (B) Fold increase in each *KRAS* mutant sequence calculated from (A) after the third round of CUT-PCR. The cut-off baseline for *KRAS* mutated ctDNA observation was determined by averaging the mutant AFs of CRISPR untreated sample in healthy controls. (Error bars mean s.e.m.; n= 3; ** $P < 0.01$, *** $P < 0.001$)

Table 4. Characteristics of tested plasma samples and comparison of cfDNA detection by CUT-PCR enrichment and canonical pyrosequencing from tumor tissue. (P-values < 0.01)

Sample No.	Type	Stage	Tumor tissue, Pyrosequencing	cfDNA, CUT-PCR
P1	CRC	IIIB	Wild-Type	c.35G>T, c.34G>T, c.34G>C
P2	CRC	I	Wild-Type	c.35G>A, c.35G>T, c.34G>T, c.34G>C
P3	CRC	IIIA	Wild-Type	c.35G>A, c.35G>T
P4	CRC	IIIB	c.35G>A	c.35G>A, c.35G>T, c.34G>T
P5	CRC	IIIC	c.35G>T	c.35G>A, c.35G>T, c.34G>C
P6	CRC	IIIB	c.35G>A	c.35G>T, c.34G>T,c.34G>C
P7	CRC	IIIB	c.38G>A†	c.35G>A, c.35G>T, c.34G>C
P8	CRC	IIA	Wild-Type	Wild-Type
H1	Healthy control	NA	NA	Wild-Type
H2	Healthy control	NA	NA	Wild-Type
H3	Healthy control	NA	NA	Wild-Type
H4	Healthy control	NA	NA	Wild-Type

Discussion

The CUT-PCR method enriches and thus enables the sensitive and precise detection of extremely small amounts of circulating mutant DNA sequences derived from tumor cells. This enrichment is realized via the removal of background signals through the specific cleavage of wild-type sequences by CRISPR endonucleases in vitro. I emphasize four advantages of CUT-PCR: First, CUT-PCR can enrich various sequences with a variety of mutation patterns, ranging from single base-pair substitutions to multi-base insertions or deletions by exploiting the strict PAM recognition property of the CRISPR system. Based on the crystal structure of the SpCas9 complex (Yamano et al. 2016), which showed that PAM recognition is critically required for cleavage of target DNA, I designed a gRNA that resulted in complete cleavage of the wild-type sequence but not mutant DNAs that lacked a proper PAM sequence. Second, CUT-PCR can enrich sequences containing a broad range of recurrent missense mutations by making use of the flexible design possibilities for CRISPR endonucleases that target wild-type DNAs. I have calculated that CUT-PCR is applicable to 80% of the oncogenic mutations registered in the COSMIC database. If researchers engineer existing CRISPR endonucleases to have altered PAM specificities44 or discover new CRISPR endonucleases that recognize different PAM sequences, CUT-PCR targetable sites would be further

extended. Third, cleaving target genomic DNA in vitro prior to PCR amplification increases the fidelity of mutant DNA enrichment. By eliminating the chance of enriching false mutations, which may be generated by DNA polymerase during PCR amplification, I can precisely enrich the true variants. Finally, by using CUT-PCR together with targeted deep sequencing, I can simultaneously enrich and identify mutant DNAs ranging from simple plasmid DNA to small amounts of cfDNAs from a patient's blood plasma.

Early detection of cancer is very important if the disease is to be treated successfully before tumors enlarge or metastasize to other organs. Statistical medical studies have shown that cancer diagnosis at an early stage generally improves patient survival rates. As an example of an application of CUT-PCR to cancer detection, I enriched extremely small amounts of sequences containing oncogenic mutations from CRC patients' blood plasma. In contrast to the results from healthy controls, CUT-PCR highly enriched circulating mutated *KRAS* sequences from cancer patients. Our results show that CUT-PCR works well over all stages of colon cancer. In addition to the *KRAS* oncogenic mutation, I argue that our system can be efficiently applied to enrich sequences containing a broad range of high ranked recurrent missense mutations, from oncogenes such as *EGFR*, *GNAQ*, or *CTNNB1*, by simply altering a component of the CRISPR system – the CRISPR endonuclease or the guide RNA. These conveniences of the CUT-PCR process could be further applied the early diagnosis of other types of cancer. I assume that CUT-PCR should continue to improve as CRISPR and

sequencing technologies evolve and could be adopted for the early and precise diagnosis of cancer.

Chapter III. Generation of *Crltd* mutants in *Chlamydomonas reinhardtii* by DNA-free CRISPR-Cas methods

Introduction

Chlamydomonas reinhardtii is unicellular free-living green algae and a well-established model organism for photosynthetic unicellular organism (Harris 2001). *C. reinhardtii* cells usually do vegetative cell growth with haploid genome (Harris 2001) and thus they are relatively easy to edit a genome than a diploid organism.

Programmable nucleases, such as zinc finger nucleases (ZFNs) (Sizova et al. 2013), transcription-activator-like effector nucleases (TALENs) have been applied to microalgae genome editing (Daboussi et al. 2014, Gao et al. 2014). CRISPR-Cas was also introduced to targeted genome editing of microalgae (Jiang et al. 2014). DNA-free CRISPR method, the delivery of Cas9-gRNA ribonucleoprotein (RNP) complex to the cells, was reported as an efficient genome editing without foreign DNA (Baek et al. 2016). Also, RNP delivery reduced off-target effects because it was rapidly degraded (Cho et al. 2014).

LTD (LHCP TRNSLOCATION DEFECT) gene encodes a chloroplast localized ankyrin protein (Ouyang et al. 2011). *Arabidopsis ltd* mutant cannot grow photoautotrophically and has yellow leaves because of mutant have reduced total chlorophyll contents than wild type (Ouyang et al. 2011). LTD protein interacts to cpSRP43 (chloroplast recognition particle 43), TIC complex and LHCP (light-harvesting chlorophyll a/b-binding proteins). It is suggested that LTD protein play an important role in LHCP

trafficking in *Arabidopsis*.

To investigate a function of LTD in *C. reinhardtii*, *Crltd* mutants were obtained through a targeted gene editing using DNA-free CRISPR method. A physiological study about *Crltd* mutant revealed that CrLTD was involved in the transport of LHCP, but not stringently required for survival in *C. reinhardtii.*

Materials and Methods

1. CRISPR-Cas9-RNP driven mutagenesis

Recombinant SpCas9 protein was expressed in *Escherichia coli*. For sgRNA construction, a partial DNA duplex were used as the mother template for *in vitro* RNA transcription. The partial duplex was annealed and then extended by Phusion polymerase (New England Biolabs) and purified with a PCR cleanup kit (DOCTOR PROTEIN, MD008). sgRNA was then generated by *in vitro* transcription at 37 ℃ overnight, after which the template DNA was cleaved with DNA nuclease I (New England Biolabs) at 37 ℃ for 30 minutes. The RNA product was further purified using a PCR cleanup kit. Cas9 protein (200 μg) and sg RNA (140 μg) were mixed and preincubated for 10 min at room temperature. Then, *Chlamydomonas* cells (5×10^5 cells) were transformed with the RNP complex in a Gene Pulser Xcell Electroporation System (BioRad) according to the protocol from the GeneArt Chlamydomonas Engineering kit (Invitrogen). After transformation, cells were incubated in TAP liquid medium with 40 mM sucrose for 12 h and harvested for genomic DNA extraction, or immediately diluted (2×10^3 cells) and plated on TAP medium containing 1.5% agar to obtain single colonies for further investigation.

2. Targeted deep sequencing

Genomic DNA segments that encompass the nuclease target

sites were amplified using Phusion polymerase (New England Biolabs). Equal amounts of the PCR amplicons were subjected to paired-end read sequencing using Illumina MiSeq. Obtained NGS data were analyzed using Cas-Analyzer. Reads that occurred only once were excluded to remove errors associated with amplification and sequencing. Indels located around the Cas9 cleavage site (3 bp upstream of the protospacer-adjacent motif sequence) were considered to be mutations induced by Cas9. To examine the potential off-target sites, I used Cas-OFFinder (Bae et al., 2014) to list potential off-target sites that differed from on-target sites by up to 4 nucleotides without a DNA or a RNA bulge in length.

Table 5. Primers used in this study.

Primer	Sequence (5' to 3')
LTDg1_off1_1stF	GCTACTTCCTGCGAATCCTG
LTDg1_off1_1stR	TCGTTTGCATTGACAAGCTC
LTDg1_off2_1stF	GCCTAGTGGGTCGTTTCAGT
LTDg1_off2_1stR	GAGGATGACGTCGGTCAAAC
LTDg1_off3_1stF	ATTGCTCTGTTGGGCCATAA
LTDg1_off3_1stR	CCTCCACCCAATCCTCTCTT
LTDg1_off4_1stF	TTACATGCCCTGTTCTGCAA
LTDg1_off4_1stR	TGTCGCTCCAGCTACGTAAA
LTDg1_off5_1stF	GCGAGCCTAGTAGCTGAACG
LTDg1_off5_1stR	ATCGACAGTGCAAGCATCAG
LTDg1_off6_1stF	GCCAGTATACCGAGCAGGAT
LTDg1_off6_1stR	GCACAATGAGCACAACGAAT
LTDg1_off7_1stF	CACGCTAACGGGGTTATGTT
LTDg1_off7_1stR	GACGCCAGTTCCAATGAAG
LTDg1_off1_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT TCACGAACTAGCGGTTGCGT
LTDg1_off1_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT TGGTATTGGAGGAGGCCTTG
LTDg1_off2_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT TTGAGCTGTCTGTCCACTGC
LTDg1_off2_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT CGCTGACCACACCTGCTCTG
LTDg1_off3_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT CGCACGGTGTCGACCCTTCC
LTDg1_off3_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT agctgcccACCATGTGGTC
LTDg1_off4_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT CCTGACGAGGCATCTGGGTA
LTDg1_off4_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT CCGATCCTAACACTGCTGAG
LTDg1_off5_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT CAAGGGGACATACCCAGGC
LTDg1_off5_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT CCGCAATGGCTTTACGAGTT
LTDg1_off6_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT CACGTCGCTGCTCAGATGC
LTDg1_off6_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTATAGAGCACGTCCGCTGTGC
LTDg1_off7_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT GCAGGTTTGTGCAAGAGTGG
LTDg1_off7_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT CAGCTGcaccgctgccagt

(continued)

Primer	Sequence (5’ to 3’)
crLTD_g1_T7	GAAATTAATACGACTCACTATAGAAGGGCGATGCGCTGCATTAGTTTTAGAGCTAGAAATAGCAAG
crLTD_g2_T7	GAAATTAATACGACTCACTATAGATGGCGCACAGTGCGAGCCAGTTTTAGAGCTAGAAATAGCAAG
crLTD_g3_T7	GAAATTAATACGACTCACTATAGTCATCTTCTCGCCATATCTCGTTTTAGAGCTAGAAATAGCAAG
crLTD_g4_T7	GAAATTAATACGACTCACTATAGAAGTTGGCAGTCACATTGCGGTTTTAGAGCTAGAAATAGCAAG
Ext_T7_rev	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC
LTD_1stF	GCCGCTCACGTTTGCAATCT
LTD_1stR	CTGCTTGCCATCAATACCCT
LTD_g12_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT CGTCCATACCAAAGCCCGGA
LTD_g12_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT AAGTTGGCAGTCACATTGCG
LTD_g34_2ndF	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT CATCCCCCACAAACTCGACC
LTD_g34_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT AGGTGTACTCATCGCGACCC

Results

1. Selection of guide RNA target sequences in *CrLTD* gene

To generate *Crltd* mutants in *Chlamydomonas reinhardtii*, guide RNA targets for SpCas9 originated from *Streptococcus pyogenes* were selected from genomic sequences of *CpLTD* gene. If Cas9 nuclease-induced indel mutations lead to frame-shift in coding sequences, they can disrupt gene function. For complete knock-out of a gene, guide RNA target sequences were chosen within a first half of coding sequences (Figure 11). Selected guide RNA targets were verified that they did not have potential off-target sites which were bearing up to 3 nt mismatches in *C. reinhardtii* genome by CAS-OFFinder (http://www.rgenome.net/cas-offinder/) (Bae et al. 2014). RNA-DNA bulges were not considered.

***CrLTD* gene**

	RGEN Target (5' to 3')	Position	Cleavage Position (%)	Direction	GC Contents (%, w/o PAM)	Out-of-frame Score	Mismatches				
							0	1	2	3	4
CrLTD_g1	AAGGGCGATGCGCTGCATTATGG	55	10.7	-	55	65.5	1	0	0	0	10
CrLTD_g2	ATGGCGCACAGTGCGAGCCACGG	97	18.1	-	65	68.9	1	0	0	0	14
CrLTD_g3	TCATCTTCTCGCCATATCTCGGG	132	26.3	+	45	60.3	1	0	0	0	1
CrLTD_g4	AAGTTGGCAGTCACATTGCGGGG	118	47.9	-	50	59.4	1	0	0	0	2

Figure 11. Description of target sequences of gRNAs used to target the *CrLTD* gene.

2. Transfection of ribonucleoproteins (RNPs) to *Chlamydomonas*

RNP (ribonucleoproteins) transfection to *Chlamydomonas* basically followed the method that was previously reported (Baek et al. 2016), then a detailed method was optimized in this work (Yu et al. 2017). Purified-SpCas9 protein and *in vitro* transcribed sgRNA were pre-mixed for formation of RNP complex. Pre-assembled RNP complex transfected to *Chlamydomonas* cells by electroporation.

To obtain single mutant colonies, cells from the transformation mixture were plated to TAP agar plate. The rest of the cells were harvested for genomic DNA extraction for use in targeted deep-sequencing analysis.

3. Analysis of CRISPR-Cas induced mutation using targeted deep-sequencing

Targeted deep-sequencing analysis were performed from pooled genomic DNA of transformants. NGS (Next generation sequencing) library was prepared by PCR-based method. Adaptor sequences were incorporated to amplicon library through nested-PCR. Prepared NGS library was sequenced by Illumina Miseq system.

NGS data from pooled *Chlamydomonas* transformants was analyzed using Cas-Analyzer (http://www.rgenome.net/cas-analyzer)(Park et al. 2017). Cas9 and CrLTD_g1 transfected

samples had 1.75% mutation rate, others were not significant (Table 6). Except CrLTD_g1, mutations were hardly observed in other samples. It suggested that target sequences of sgRNA affected Cas9-sgRNA complex activity in *Chlamydomonas*.

Through more detailed sequence analysis, the most abundant mutation patterns were found in CrLTD_g1 treated sample (Table 7). 1 bp insertion was a dominant mutation pattern and different types of nucleotide were incorporated in mutated sequences.

Table 6. The frequency of mutations (insertion and deletions; indels) in the wild-type (control) and RNP-transfected cells (experiment) measured for each gRNA by targeted deep sequencing.

	Target sequences	Sample	total count	mutated count	mutated ratio(%)
crLTD_g1	AAGGGCGATGCGCTGCATTATGG	control	55401	9	0.02%
		experiment	52027	913	1.75%
crLTD_g2	ATGGCGCACAGTGCGAGCCACGG	control	55401	9	0.02%
		experiment	50897	9	0.02%
crLTD_g3	TCATCTTCTCGCCATATCTCGGG	control	61508	2	0.00%
		experiment	53492	9	0.02%
crLTD_g4	AAGTTGGCAGTCACATTGCGGGG	control	61508	2	0.00%
		experiment	46198	6	0.01%

Table 7. Mutation pattern analysis of RNP-transfected cells for CrLTD_g1. The top three indel patterns are listed.

Top Rank	Sequences	Insertion /Deletion	Mutation Counts
1	ACTCGACCATAA-TGCAGCGCATCGCCCTTGCCTCG - WT ACTCGACCATAATTGCAGCGCATCGCCCTTGCCTCG	+1	222 counts
2	ACTCGACCATAA-TGCAGCGCATCGCCCTTGCCTCG - WT ACTCGACCATAAATGCAGCGCATCGCCCTTGCCTCG	+1	206 counts
3	ACTCGACCATAA-TGCAGCGCATCGCCCTTGCCTCG - WT ACTCGACCATAACTGCAGCGCATCGCCCTTGCCTCG	+1	123 counts

4. Isolation of *Crltd* mutants

A previous study reported that LTD protein was essential for the import of LHCP (light-harvesting chlorophyll a/b-binding proteins) to thylakoid membrane in *Arabidopsis* (Ouyang et al. 2011). *Arabidopsis ltd* mutant had yellow leaves due to reduced chlorophyll contents than wild-type. As phenotype of *Arabidopsis ltd* mutant and predicted function of CrLTD protein, it was predicted that *Crltd* mutant has pale-green color than wild-type.

Among colonies of *Chlamydomonas* transformants, pale-green colonies were selected and confirmed by Sanger sequencing. Four strains of *CrLTD* knock-out mutants were isolated from CrLTD_g1 transformants (Figure 12). Interestingly, two mutants were identical sequences which were identified as top rank in the targeted deep-sequencing. *Crltd1* mutant had the first-ranked mutation pattern and *Crltd4* had the fourth-ranked pattern (Figure 12 and Table 7).

Figure 12. Sanger sequencing chromatograms for CRISPR-Cas9 induced *Crltd* mutant strains. DNA sequences of four *Crltd* strains were confirmed by Sanger sequencing. The start codon of the wild type is indicated with a red box. Ins, insertion; del, deletions.

5. Phenotypes of *Crltd* mutants

The *Crltd* mutants had pale green color since it had a lower total chlorophyll content than the wild type. Likewise, the levels of the core proteins of PSI were reduced than their wild-type levels. Overall structure of chloroplast and cellular organization was changed in *Crltd* mutant. The number of thylakoid membrane layers was increased in the mutant chloroplast and abnormally enlarged vacuoles were observed in the *Crltd1* mutant (Figure 13).

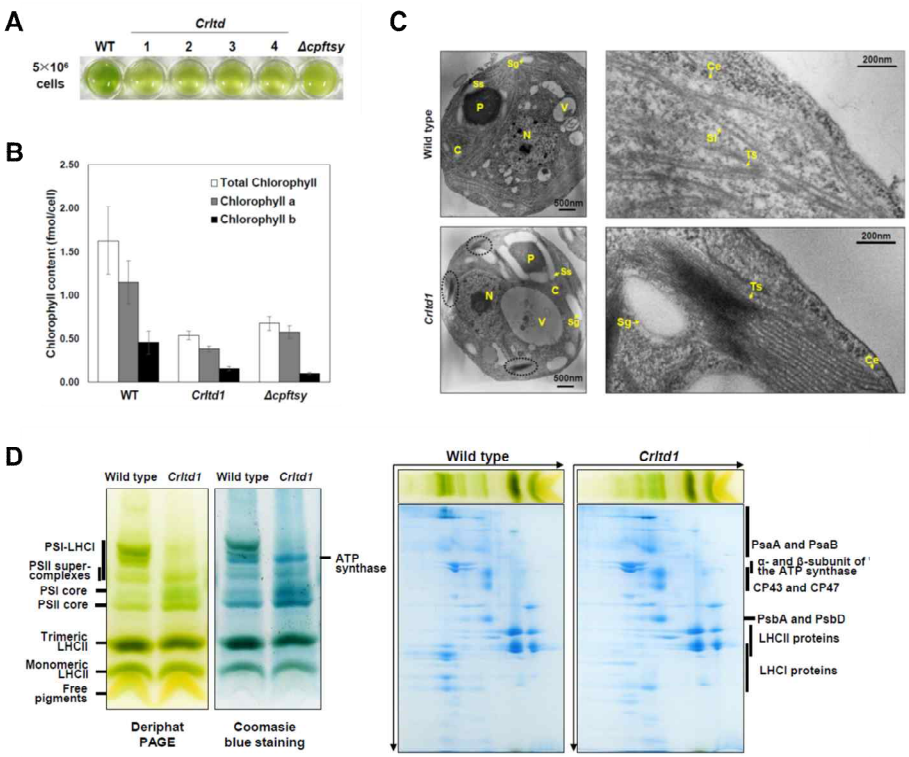


Figure 13. Phenotypes of *Crltd* mutants. (A) Coloration of *C. reinhardtii* wild type and *Crltd* strains. (B) Total chlorophyll content of the wild type and the *Crltd1* and *⊿cpftsy* strains. (C) Transmission electron microscopy images of the cross-sections from wild-type (upper) and *Crltd1* (bottom). N, Nucleus; V, Vacuole; C, Chloroplast; Ce, Chloroplast envelope; P, Pyrenoid; Ss, Starch sheath; Sg, Starch granule; Ts, Thylakoid stacks; Sl, Stroma lamellae. (D) Deriphat-PAGE and 2-dimensional PAGE analysis of thylakoid membrane protein complexes in wild-type and *Crltd1* mutant.

6. Off-target analysis in *Crltd* mutant

A targeting of CRISPR is based on the complement sequence hybridization between DNA and guide RNA (Jinek et al. 2012, Nishimasu et al. 2014). Accordingly, DNA-RNA hybridization can occur with bearing some of mismatched nucleotides. These off-target mutations could cause unwanted mutations.

To inspect off-target mutations in *Crltd* mutants, Potential off-targets which had up to 4 nucleotide mismatches to CrLTD_g1 target sequences were selected using Cas-OFFinder in the genome of *Chlamydomonas reinhardtii* (V5.0). 7 off-target loci in a genome were analyzed by targeted deep-sequencing in wild type and *Crltd1* mutant cells. As a result, mutations in off-target loci were not detected in *Crltd1* mutants and only detected below error rates in wild type (Table 8).

It supposed that a usage of Cas9-sgRNA ribonucleoprotein could reduce off-target effects in *Chlamydomonas*. Cas9-sgRNA ribonucleoprotein can act much faster than a plasmid expressing Cas9, but also degrades rapidly. A short duration of Cas9-gRNA complex reduced off-target mutations (Cho et al. 2014).

Table 8. Analysis of off-target effects in the wild type and *Crltd1*. Mutation frequencies at potential off-target sites of the CrLTD gene-specific CrLTD_g1 were measured by targeted deep sequencing in the wild type and *Crltd1*. Potential off-target sites that differed from on-target sites by up to 4 nucleotides were selected. Different nucleotides between on-target and off-target were highlighted in Red.

Target (5′ to 3′)	Wild type			*Crltd1*		
	Total reads	Indel	Indel frequency (%)	Total reads	Indel	Indel frequency (%)
gAGGcCGATGCGCTaCgTTATGG	14321	0	0.0	8584	0	0.0
AAcGGCcAacCGCTGCATTACGG	27539	0	0.0	20148	0	0.0
AAGGGCGAgGCGCTGCtggAGGG	31882	2	0.006	21134	0	0.0
cAGGaCcATGCGCTGCAcTAGGG	37847	0	0.0	30822	0	0.0
AAGGGCGcTGtGCcGCAaTACGG	40377	0	0.0	27765	0	0.0
AAcGGCGcTGCGCTGCATcgTGG	29388	0	0.0	17057	0	0.0
AAcGGCGgcGCGCTGCAgTACGG	31603	0	0.0	10426	0	0.0

Discussion

In this study, knock-out mutants of CrLTD gene were generated by RNP delivery in *Chlamydomonas*. Though LTD mutants were reported in *Arabidopsis* (Ouyang et al. 2011), *CrLTD* mutants were reported at first in *Chlamydomonas* from this study. It was helpful for investigating CrLTD function in *Chlamydomonas.* In technically, an optimization of DNA-free CRISPR method made easy the generation of knock-out mutant in *Chlamydomonas* (Yu et al. 2017).

Because DNA-free CRISPR method did not introduce foreign DNA to the cells, it could avoid GMO issues when mutants would be used to a material for a food, cosmetics and drug. In addition, off-target analysis in *Crltd1* mutant revealed that unwanted mutations did not occur by this method as previously known that (Cho et al. 2014, Baek et al. 2016).

DNA-free CRISPR method is an efficient tool to generate knock-out mutant in *Chlamydomonas*. Moreover, it could be applied to the engineering of other industrial microalgae. Microalgae including *Chlamydomonas* are important living-cell factories for production of bio-fuels and biochemicals used in food, drug, cosmetic industries (Fu et al. 2016). Therefore, a genome engineering is promising for increasing the productivity of microalgae.

Chapter IV. Targeted demethylation using dCas9 fused *Arabidopsis* DME protein

Introduction

A methylation of cytosine (5mC) is an epigenetic marker conserved in plant, animal and fungi (Law et al. 2010, Wu et al. 2014). DNA methylation has crucial effects on genome stability, gene expression and development.

De novo DNA methylation is established by DNMT3A and DNMT3B in mammal, DRM1 and DRM2 in plant. During cell division, DNA methylation is maintained in daughter cells by DNMT1 in mammal, MET1 and CMT3 in plant (Law et al. 2010).

Loss of DNA methylation can occur passively during replication. But the regulated, active demethylation pathway is important for a developmental process and gene activation (Wu et al. 2014). In mammal, DNA dioxygenase TET proteins convert 5-methylcytosine (5mC) to 5-hydroymethylcytosine (5hmC). Further TET proteins oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). End products of TET proteins, 5fC and 5caC are removed by uracil DNA glycosylase TDG. Excised based is repaired by BER (base excision repair) pathway and non-methyl cytosine is incorporated at the abasic site. As a result, a methylcytosine is replaced to a cytosine. Unlike mammal system, plant DNA glycosylase DME family (DME, DML1, DML2, DML3) recognizes 5mC and directly releases 5mC base from DNA. Following BER repair system is identical to mammal system.

DME (*DEMETER*) gene encodes a DNA glycosylase domain protein and DME protein functions as an active DNA demethylase in

plants (Choi et al. 2002, Gehring et al. 2006). DME regulates *MEA* (*MEDEA*) gene expression through demethylation at a promoter region and 3' repeat region (MEA-ISR) of *MEA* in an endosperm.

DME protein has three domains; Domain A, Glycosylase domain and Domain B. Glycosylation domain is catalytically active domain. Though biochemical functions of domain A and domain B have not been identified, they are required for a demethylase activity. These three domains are essential for demethylation activity of DME protein, it solely can remove methylcytosine *in vitro* (Mok et al. 2010).

Targeted demethylation in human cells was previously reported. ZF fused human TET1 enzyme was used to induce demethylation in human *ICAM1* promoter (Chen et al. 2014), TALE-TET1 fusion protein could activate human *KLF4* gene through demethylation of CpG island (Maeder et al. 2013). Inactive Cas9 fused TET1 protein was also applied to modify DNA methylation moiety in human cells (Choudhury et al. 2016, Liu et al. 2016, Xu et al. 2016). Even though considerable studies on targeted demethylation using TET1 protein, it could not be representative. Because TET1 fusion protein has low efficiency and the recovery of DNA methylation were observed in following days (Maeder et al. 2013).

According to different mechanisms of DNA demethylation between plant and human, especially different enzymatic activities between DME and TET1 protein, I designed a novel targeted demethylation tool. Catalytic core domain of DME which consisted of domain A, glycosylation domain and domain B were fused to

inactive Cas9. Unfortunately, I could not observe targeted demethylation effects by dCas9-DME in plant and human cells, it still had a probability to become an attractive tool for epi-genome editing in plants.

Materials and Methods

1. Cell culture and transfection

K562 (ATCC, CCL-243) cells were maintained in RPMI 1640 medium supplemented with 10% FBS and 1% antibiotics. For electroporation, $2x10^5$ cells were transfected with Cas9-DME expression plasmids (500 ng) and sgRNA plasmids (500 ng using 4D-nucleofector (Lonza) in 20 μl Nucleovette Strips. Genomic DNA was isolated with the Nucleospin Tissue Kit (MACHEREY-NAGEL) 72 h post-transfection.

2. Bisulfite conversion

Bisulfite conversion was performed according to the instructor's manual. EZ DNA Methylation-Direct Kit (Catalogue No. D5020, Zymo Research, Inc., Irvine, CA, USA) were applied. For bisulfite reaction, 500 ng of genomic DNA were used in human cell line experiments and 100 ng of genomic DNA were used in plant transgenic experiments.

3. Targeted deep-sequencing

Target sites were amplified with adaptor primers using Phusion polymerase (New England Biolabs). The resulting deep sequencing libraries were subjected to paired-end sequencing with the MiniSeq system (Illumina). After MiniSeq, paired-end reads were joined by the Fastq-join.

4. Analysis of DNA methylation

Analysis tool is python. Python gained the reference sequence which ranges from −70 base to 70 base for cleavage site of SpCas9 and the indicator sequence which is defined as both end of the reference sequence. Using the indicator sequence, in fastqjoin file, valid sequences are collected by containing both indicators with up to a 1 nucleotide mismatch. Then python counts the frequency of each sequence. Using EMBOSS needle, python aligns the counted sequences with reference sequences. Reads which have equal length with the reference sequence were sorted. A proportion of nucleotides at each positions were analyzed.

Table 9. Primers used in this study.

Primer	Sequence (5' to 3')
BspEI_3XGGS DME F	TCCGGAGGTGGAAGTGGCGGCAGCGGAGGTTCTTACAAAGGAGATGGTGCACTTGT
DMEcore_AInk	GCCTGATCCGCCGCTACCCCCAGAACCTCCTGGAGGGAATCGAGCAGCTA
DMEcore_InkG	GGGGGTAGCGGCGGATCAGGCGGGAGTACAAGCCAGTGGGATAGTCT
DME_BspEI R	TCCGGAGGTTTTGTTGTTCTTCAATTTGC
BspEI_6XGGS	TCCGGAGGTGGAAGTGGCGGCAGCGGAGGTTCTGGTGGAAGTGGCGGCAGC
dCAS9 D10A F	GCGGCCGCACTAGTGATATC
dCAS9 D10A R	GTTGGTACCGATGGCCAGGCCGATGC
dCAS9 H840A 1F	GAGCCGCAAGCTTATCAACGGCA
dCAS9 H840A 1R	TGGGGCACGATGGCGTCCACGTCGTAGTC
dCAS9 H840A 2F	TACGACGTGGACGCCATCGTGCCCCAGAG
dCAS9 H840A 2R	GCTCAAGCTAAGCTTGGGCCCG
KLF4_2ndF1_BSn	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTGTTTYGAATAAYGTGAGTAT
KLF4_2ndR1_BSn	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAAACCCRACTTACRCCCCA
KLF4_2ndF2_BSn	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGTTTTTGGTTTAGTTGTGTATG
KLF4_2ndF3_BSn	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGTTTTTATTAGTTTTYGGG
KLF4_2ndR3_BSn	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATTATATCCTATCTACCCAATTAC
KLF4_2ndF1_wtn	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTGCCCCGAATAACGTGA
KLF4_2ndR1_wtn	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGACCCGGCTTGCGCCC
KLF4_2ndF2_wtn	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGTCCTTGGCTCAGCTGTG
KLF4_2ndF3_wtn	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGTCTTCACCAGTCCCCG
KLF4_2ndR3_wtn	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTTATGTCCTGTCTGCCCAA

(continued)

Primer	Sequence (5’ to 3’)
KLF4sg1_up	CACCG AGAGCCACTGAACGAGGGTA
KLF4sg1_down	AAAC TACCCTCGTTCAGTGGCTCT C
KLF4sg2_up	CACCG ACACGGAAGCTATCCCGGGA
KLF4sg2_down	AAAC TCCCGGGATAGCTTCCGTGT C
KLF4sg3_up	CACCG CCGGGTCTTCACCAGTCCCC
KLF4sg3_down	AAAC GGGGACTGGTGAAGACCCGG C
KLF4sg4_up	CACCG CCAGAGTGAGTTTAGCGCGT
KLF4sg4_down	AAAC ACGCGCTAAACTCACTCTGG C
KLF4sg5_up	CACCG TCCCCGGAATTGGCACACCG
KLF4sg5_down	AAAC CGGTGTGCCAATTCCGGGGA C
FWAsg1_up	GATTG TTCTCGACGGAAAGATGTAT
FWAsg1_down	AAAC ATACATCTTTCCGTCGAGAA C
FWAsg2_up	GATTG TATGAATGTTGAATGGGATA
FWAsg2_down	AAAC TATCCCATTCAACATTCATA C
FWAsg3_up	GATTG CGCTCGTATGAATGTTGAAT
FWAsg3_down	AAAC ATTCAACATTCATACGAGCG C
FWAsg4_up	GATTG GAAAAGCAAAAACCCTAGAG
FWAsg4_down	AAAC CTCTAGGGTTTTTGCTTTTC C
FWAsg5_up	GATTG ATGTTGAATGGGATAAAGAG
FWAsg5_down	AAAC CTCTTTATCCCATTCAACAT C
FWA_wt_1stF1n	cttcgttcttgtgtcatgtaatag
FWA_BS_1stF1n	TttYgttTttgtgtTatgtaatagattaT
FWA_wt_1stR1n	tctagtgtctcgacaacgaac
FWA_BS_1stR1n	tctaAtAtctcRacaacRaacaaAacaa
FWA_wt_2ndF1	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT gggtttagtgtttacttgtttaagg
FWA_wt_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT aagcaaccttaaacaaccaaatagc
FWA_wt_2ndF2	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT gatttatcgatttgtgggatactg
FWA_BS_2ndF1	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT gggtttagtgtttaTttgtttaaggt
FWA_BS_2ndR	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT aaAcaaccttaaacaaccaaataAca
FWA_BS_2ndF2	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT gatttatYgatttgtgggataTtgaT

Results

1. Plasmid construction of human cell expression vectors

To express deadCas9 fused DME protein in human cell line, fusion constructs were generated. As previous report, catalytic domain of DME protein was used. DME⊿N677 was a N-terminal deletion form and DMEcore was a shorter form in which the region between Domain A and Glycosylase domain were exchanged to liker sequences additionally (Figure 14A). They are known to have demethylase activity *in vitro* (Mok et al. 2010). Two forms of DME catalytic domain were inserted to c-terminal of inactive Cas9 (dCas9, D10A, H840A mutant) which was driven by CMV (Cytomegalovirus) promoter. For proper functions of fusion protein, dCas9 and catalytic domain of DME were linked by 3xGGS (Gly-Gly-Ser) or 6xGGS (Figure 14B).

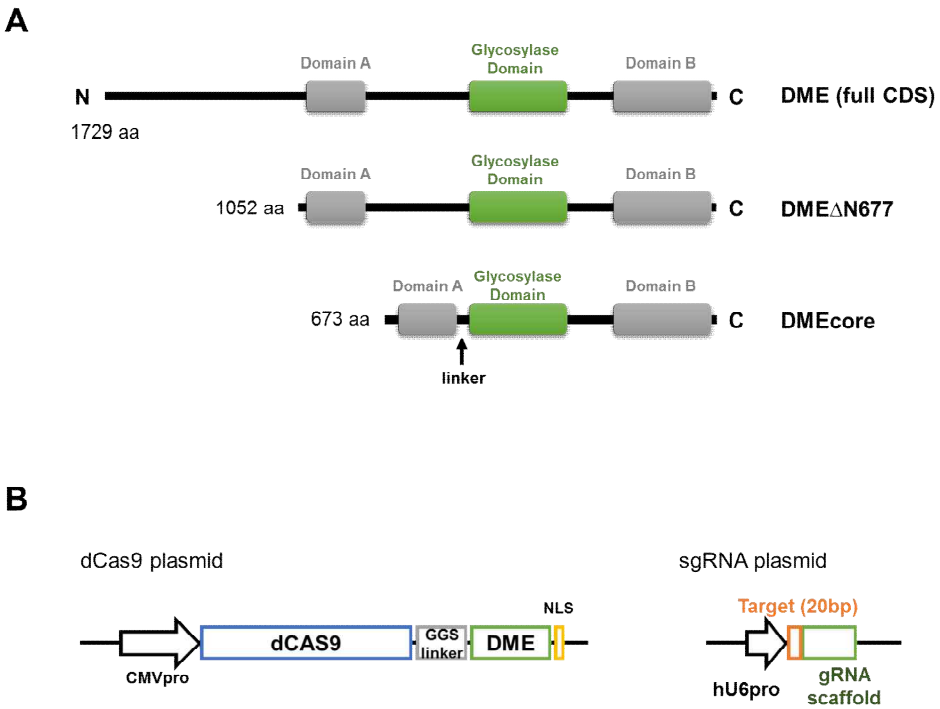
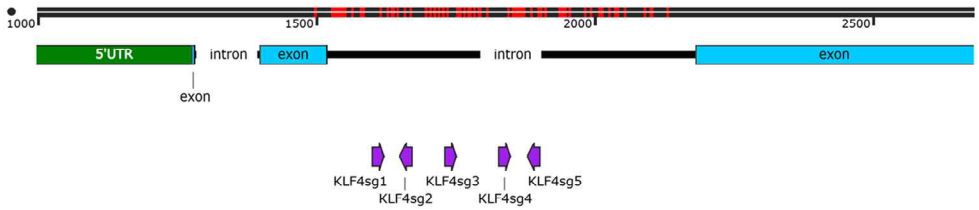


Figure 14. Diagram of dCas9-DME constructions. (A) DME protein domain and catalytic domain which were used in constructions. (B) Structure of dCas9-DME construct for the expression in human cell.

2. Targets selection in CpG island of *KLF4*

To verify targeted demethylation by dCas9-DME, *KLF4* CpG island in intron were selected because it was densely methylated in K562 cell line (Human myelogenous leukemia) and also had been targeted by TALE-TET1 for the targeted demethylation (Maeder et al. 2013).

gRNA targets were selected in the second intron of *KLF4* gene (Figure 15). Wild type SpCas9 and target sgRNA plasmid were transfected to K562 cells to compare the activity of Cas9 at each target site. Targeted deep-sequencing analysis revealed that all the selected gRNAs targeting *KLF4* had similar activity in K562 cells (Table 10).



	RGEN Target (5' to 3')	Position	Cleavage Position (%)	Direction	GC Contents (%, w/o PAM)	Out-of-frame Score	Mismatches		
							0	1	2
KLF4sg1	AGAGCCACTGAACGAGGGTAAGG	83	15	+	55	55.7	1	0	0
KLF4sg2	ACACGGAAGCTATCCCGGGAAGG	131	20.6	-	60	85.1	1	0	0
KLF4sg3	CCGGGTCTTCACCAGTCCCCGGG	213	34.6	+	70	59.1	1	0	1
KLF4sg4	CCAGAGTGAGTTTAGCGCGTAGG	308	49	+	55	73	1	0	1
KLF4sg5	TCCCCGGAATTGGCACACCGAGG	361	55.4	-	65	74.6	1	0	0

Figure 15. gRNA target selection in *KLF4* CpG island. An upper diagram displayed a partial genomic structure of *KLF4* gene. Red indicated CpG island positions and violet arrows indicated gRNA target sequences. A bottom table represented properties of selected gRNA targets.

Table 10. Mutation frequencies of *KLF4* targets using wild-type SpCas9 in human K562 cells

	target	Sample	Total count	Mutated count	Mutation frequency(%)
KLF4_g1	AGAGCCACTGAACGAGGGTAAGG	control	37262	9	0.02%
		sg1	48099	26027	54.11%
KLF4_g2	ACACGGAAGCTATCCCGGGAAGG	control	37262	9	0.02%
		sg2	32187	19879	61.76%
KLF4_g3	CCGGGTCTTCACCAGTCCCCGGG	control	37620	209	0.56%
		sg3	33605	17919	53.32%
KLF4_g4	CCAGAGTGAGTTTAGCGCGTAGG	control	38795	0	0.00%
		sg4	41526	19591	47.18%
KLF4_g5	TCCCCGGAATTGGCACACCGAGG	control	38795	0	0.00%
		sg5	34503	20104	58.27%

3. Targeted demethylation in *KLF4* CpG island

dCas9-DME plasmid and sgRNA plasmid were transfected to K562 cells by an electroporation. 3 days after the transfection, genomic DNA were purified and DNA methylation level of each sample was analyzed through a bisulfite deep-sequencing.

From NGS data, reads which had an equal sequences length were picked out. Then nucleotide compositions were analyzed in each nucleotide positions. Rates of C to T conversion by bisulfite treatment were 95-98% (Figure 16).

DNA methylation in CpG sites could be detected at a high resolution through this method. Read counts of each sample were from 30000 to 80000. Unfortunately, any significant changes of DNA methylation were not observed in most samples (Figure 17). It was assumed that because a mechanism of DNA methylation in mammalian cell (Wu et al. 2014), DME protein could not work in human cells.

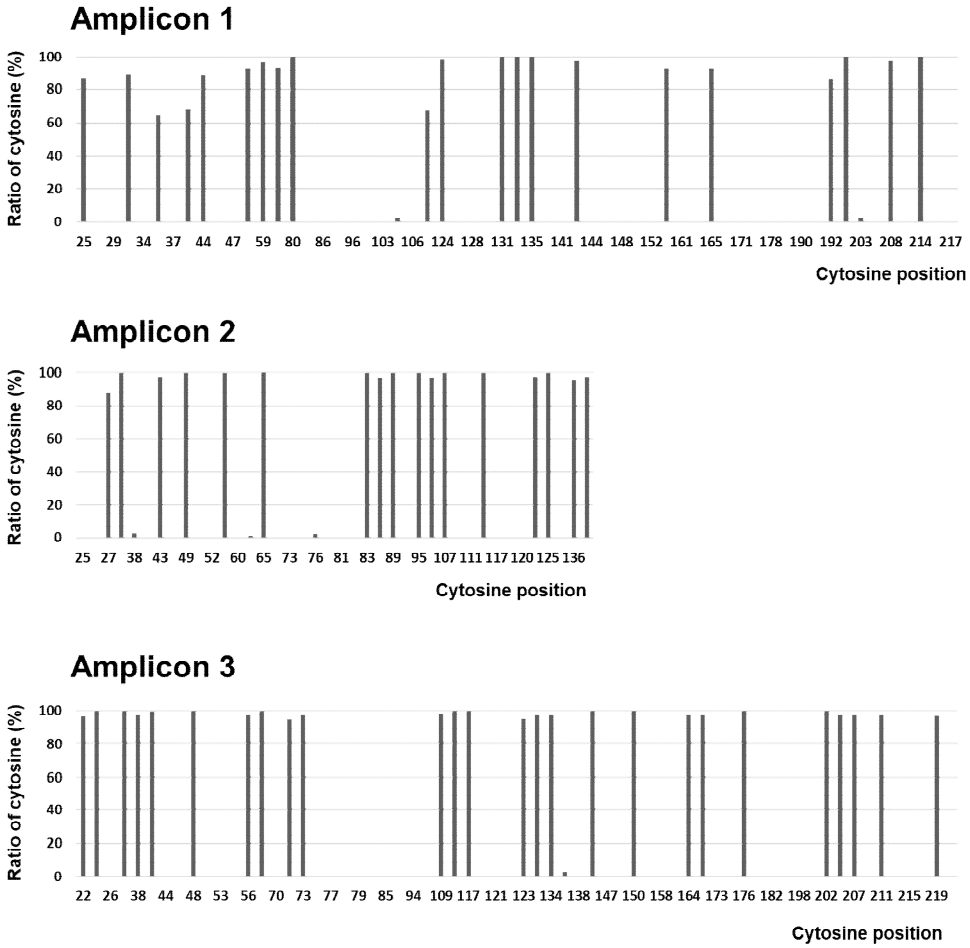


Figure 16. Bisulfite conversion in *KLF4* CpG island. Ratio of cytosine was analyzed by a deep-sequencing after bisulfite treatment in control genomic DNA. Upper; amplicon 1 which possessed KLF4_g1 and KLF4_g2 targeting regions. Middle; amplicon 2 which possessed KLF4_g3 targeting region. Bottom; amplicon 3 which possessed KLF4_g4 and KLF_g5 targeting regions.

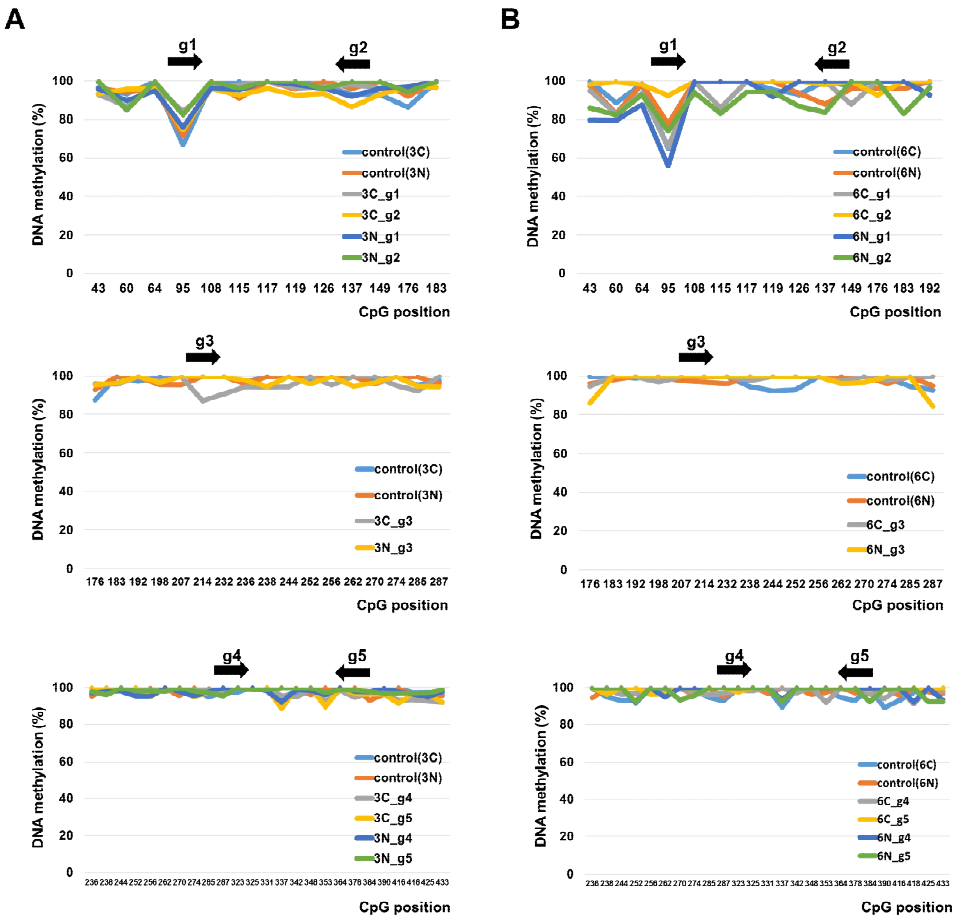


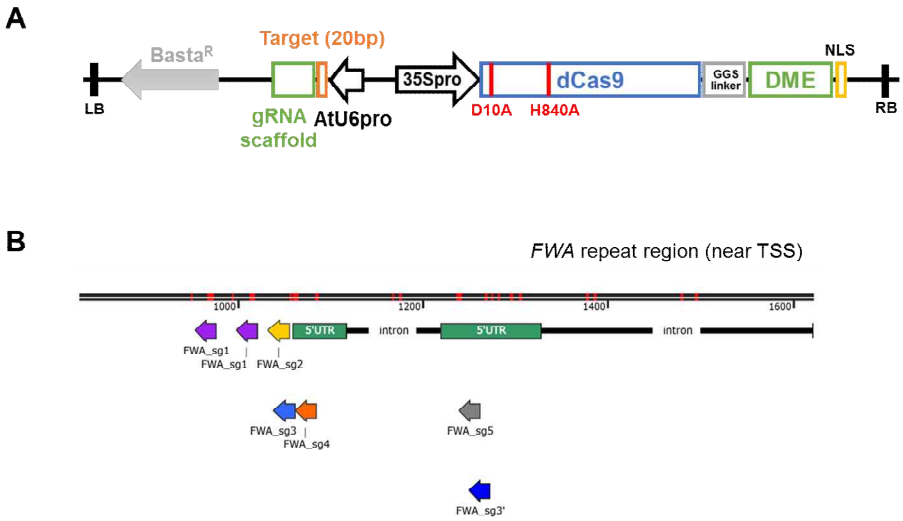
Figure 17. DNA methylation levels in *KLF4* CpG island. DNA methylation levels at *KLF4* 2nd intron was analyzed by a bisulfite deep-sequencing after the transfection of dCas9-DME and sgRNA. (A) A Short linker (3xGGS) fusion protein (3xGGS-DMEcore, 3xGGS-DME⊿N677) used samples. (B) A longer linker (6xGGS) fusion protein (6xGGS-DMEcore, 6xGGS-DME⊿N677) used samples. Black arrow represented the position of each sgRNA used.

4. Generation of *fwa* epi-mutant

To investigate the targeted demethylation by dCas9-DME in Arabidopsis, *FWA* gene was chosen for a target gene. *FWA* gene had a repeat region at its promoter, DNA methylation in the repeat region regulates an expression of *FWA* (Soppe et al. 2000, Kinoshita et al. 2007). I intended to generate epi-mutants of *FWA* through the targeted demethylation at the *FWA* promoter.

The mutations for an inactivation of Cas9, D10A and H840A were introduced to plant CRISPR vector pBatC, then DME catalytic domain linked to c-terminal of dCas9 like as constructs for human expression plasmids (Figure 18A).

gRNA target sequences were selected near the transcription start site of *FWA*. It was reported that a level of DNA methylation in a TSS of *FWA* was critical for a regulation of *FWA* expression (Kinoshita et al. 2007). Since targets were selected on repeat sequences, some of gRNA could target two sites in the promoter with or without bearing mismatches (Figure 18B). A dual targeting of dCas9-DME was expected to increase an efficiency of demethylation.



	Targeting	match	RGEN Target (5' to 3')	Position	Direction	GC Contents (%,w/o PAM)	Mismatches			near CpG	targeting
							0	1	2		
FWA sg1	double	perfect	TTCTCGACGGAAAGATGTATGGG	31	-	40	2	0	0	1	2
FWA sg2	single		TATGAATGTTGAATGGGATAAGG	110	-	30	1	0	0	3	1
FWA sg3	double	1 mismatch	CGCTCGTATGAATGTTGAATGGG	116	-	40	1	1	0	1	2(1)
FWA sg4	single		GAAAAGCAAAAACCCTAGAGCGG	139	-	40	1	0	1	2	1
FWA sg5	single		ATGTTGAATGGGATAAAGAGCGG	316	-	35	1	0	0	2	1

Figure 18. A Schematic design of the targeted demethylation at a *FWA* promoter region. (A) Construct of pBAtC-dCas9-DME all-in-one plasmid for *Arabidopsis*. (B) gRNA targets in a *FWA* repeats region. Red indicated positions of methylated cytosine and arrows represented gRNA targets. Properties of gRNA target sequences were described in the table.

5. Phenotype of dCas9-DME transgenic plants

A dCas9-DME binary vector was introduced to wild-type *Arabidopsis* Col-0 plants by *Agrobacterium*-mediated transformation. Phenotypes of T1 transgenic plants which were selected by basta resistance were observed.

As the *fwa* mutant flowers lately than wild type (Soppe et al. 2000, Ikeda et al. 2007), I focused on late flowering phenotype of transgenic plants. Among T1 plants, only 1 plant (6C_g3-16, 6xGGS-DMEcore with g3) had late flowering phenotype, most of all showed a similar flowering time to wild type. In addition, an axillary vegetative meristem was observed in some of transgenic plants (Table 11). Two kinds of phenotype were only found in gRNA targeting transgenic plants (FWA_g1-5), not in non-targeting transgenic plants (-sg). These abnormal plants were considered as epi-mutant candidates, so DNA methylation analysis was performed.

Table 11. A phenotype analysis of *FWA*-targeted dCas9-DME transgenic plants in T1 generation

Constructs	gRNA	Number of T1 plants (Basta^R)	Number of plants	
			Late flowering	Axillary vegetative meristem
dCas9-3xGGS-DMEcore	(-sg), non-targeting	20	0	0
	FWA_g1	20	0	0
	FWA_g2	20	0	0
	FWA_g3	20	0	0
	FWA_g4	20	0	0
	FWA_g5	20	0	1
dCas9-6xGGS-DMEcore	(-sg), non-targeting	20	0	0
	FWA_g1	20	0	0
	FWA_g2	20	0	2
	FWA_g3	20	1	0
	FWA_g4	20	0	0
	FWA_g5	20	0	0
dCas9-3xGGS-DMEΔN677	(-sg), non-targeting	20	0	0
	FWA_g1	20	0	0
	FWA_g2	20	0	0
	FWA_g3	20	0	0
	FWA_g4	20	0	2
	FWA_g5	20	0	0
Col-0		20	0	0

6. Analysis of DNA methylation in *epi*-mutant candidates

A bisulfite deep-sequencing was performed in mutant candidates which were selected by phenotypes to confirm a level of DNA methylation at *FWA* promoter. Two regions which covered target sites in *FWA* promoter were selected (Figure 19A). The efficiency of C to T conversion by bisulfite treatment were 95-98% (Figure 19B).

There was no difference of DNA methylation level between control Col-0 and candidate transgenic plants (Figure 19C). Even though 6C_g3-16 plants showed late flowering, DNA methylation were not different to wild type. Other candidates also had similar level of DNA methylation to wild type.

Late flowering and axillary vegetative meristem phenotypes disappeared in T2 transgenic plants. It suggested that heritable changes of DNA methylation did not occurred in these plants. It was assumed that phenotype of plants might be due to an ectopic expression of dCas9-DME.

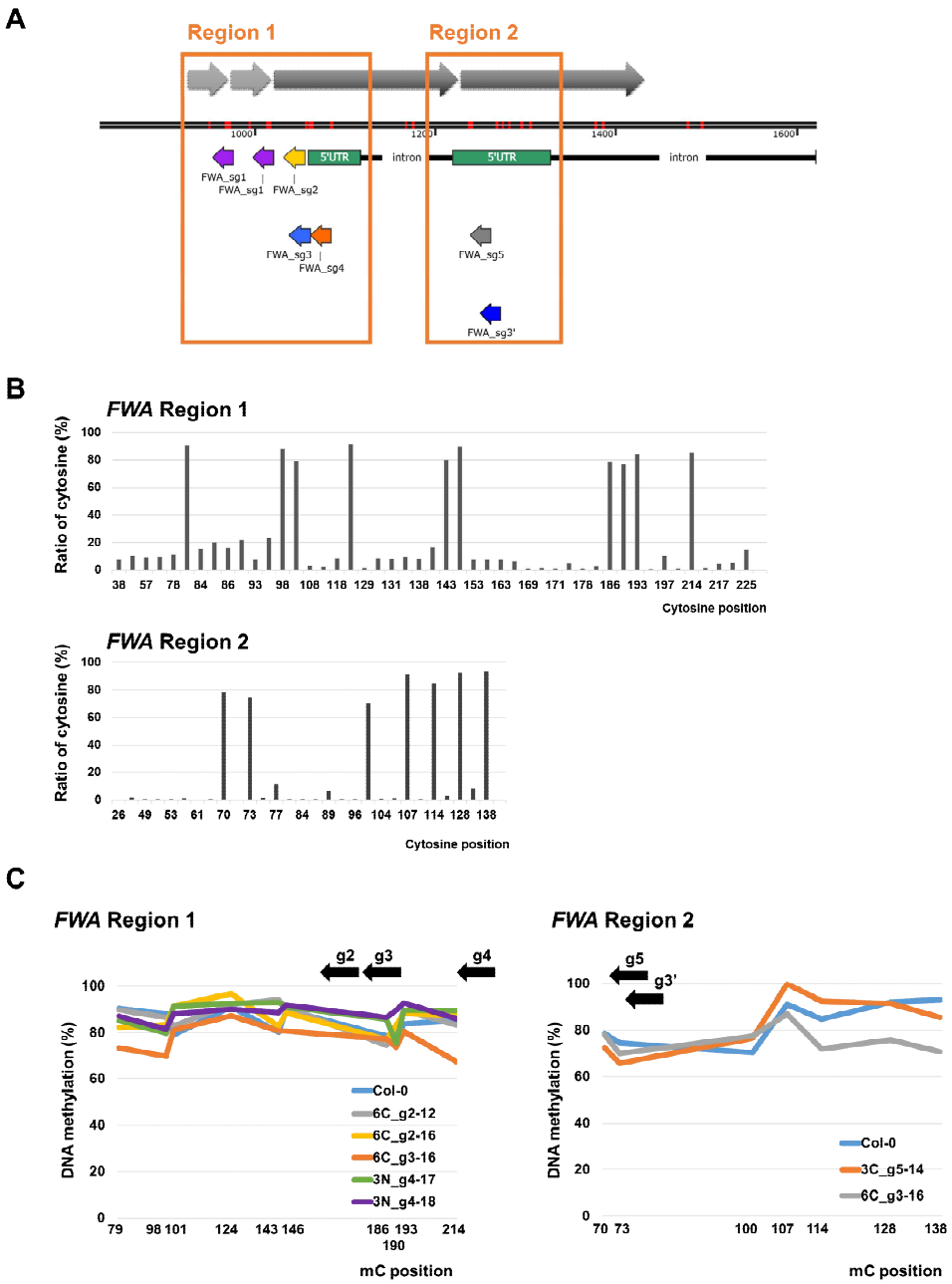


Figure 19. DNA methylation levels at *FWA* promoter by a bisulfite deep-sequencing in T1 transgenic plants. (A) Genomic structure of *FWA* promoter. Gray arrows indicated repeat sequences and orange box represented the regions which were analyzed by bisulfite

deep-sequencing. (B) Bisulfite conversion in *FWA* promoter region. (C) A level of DNA methylation in *FWA* promoter. Arrows represented gRNA targets.

7. The improvement of dCas9-DME

Based on the result above, it was suggested that dCas9-DME does not have DNA demethylation activity in a programmable manner. To confirm a DNA demethylase activity of dCas9-DME, *in vitro* activity assay will be performed using a recombinant protein which is purified from *E.coli*.

In addition, a fusion protein design for the proper activity should be optimized. DME fusion in opposite site (N-terminal of Cas9) and use of more various linker (length and property) are possible attempts to make a functional fusion protein.

Discussion

Plants have a unique active DNA demethylation mechanism in comparison with other organisms (Wu et al. 2014). DME protein which has DNA glycosylase activity is a plant-specific DNA demethylase (Choi et al. 2002, Gehring et al. 2006).

In this study, I tried the targeted demethylation using inactive Cas9 fused DME catalytic domain in human and plants. Since DME protein could remove methylated cytosine directly by its glycosylase activity (Mok et al. 2010), I anticipated more efficient demethylation activity than a human TET1 system that was previously reported (Maeder et al. 2013, Chen et al. 2014, Choudhury et al. 2016, Liu et al. 2016, Xu et al. 2016).

However, dCas9-DME could not carry out the DNA demethylation at targeted regions in human cells and even *Arabidopsis* transgenic plants (Figure 17 and 19). In *Arabidopsis* transgenic plants, some of transgenic plants which had FWA-targeting sgRNA showed a late flowering or an axillary vegetative phenotype but they were not maintained in a next generation. A stable epi-mutant had not been obtained in this study.

To improve the demethylation tool using DME, more scrupulous confirmations should be needed. First of all, the activity of dCas9-DME should be confirmed *in vitro*. From *in vitro* test, a design of fusion protein would be optimized for the activity.

As mentioned earlier, plants have a different DNA

demethylation system to other organisms. Plants do not have a methyl-cytosine modification enzyme like as human TET1 (Wu et al. 2014). Therefore, it could be assumed that dCas9-TET1 system did not work well in a plant cell. If dCas9-DME demethylation system is advanced efficiently, it would be an attractive tool in a plant epi-genome editing.

References

Alix-Panabieres, C., H. Schwarzenbach and K. Pantel. 2012. Circulating tumor cells and circulating tumor DNA. *Annu Rev Med.* **63**: 199-215.

Anders, C., O. Niewoehner, A. Duerst and M. Jinek. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature.* **513**: 569.

Anders, C., O. Niewoehner, A. Duerst and M. Jinek. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature.* **513**: 569-573.

Bae, S., J. Park and J. S. Kim. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics.* **30**(10): 1473-1475.

Baek, K., D. H. Kim, J. Jeong, S. J. Sim, A. Melis and J. S. Kim. 2016. DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. *Sci Rep.* **6**: 30620.

Bettegowda, C., M. Sausen, R. J. Leary, I. Kinde, Y. Wang and N. Agrawal. 2014. Detection of circulating tumor DNA in early-

and late-stage human malignancies. *Sci Transl Med.* 6: 224ra224.

Boch, J., H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt and U. Bonas. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science.* 326(5959): 1509-1512.

Ceccaldi, R., B. Rondinelli and A. D. D' Andrea. 2015. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology.* 26(1): 52-64.

Chan, K. C., P. Jiang, Y. W. Zheng, G. J. Liao, H. Sun and J. Wong. 2013. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem.* 59: 211-224.

Chang-Hao Tsao, S., J. Weiss, C. Hudson, C. Christophi, J. Cebon and A. Behren. 2015. Monitoring response to therapy in melanoma by quantifying circulating tumour DNA with droplet digital PCR for BRAF and NRAS mutations. *Sci Rep.* 5: 11198.

Chang, N., C. Sun, L. Gao, D. Zhu, X. Xu and X. Zhu. 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* 23: 465-472.

Chen, H., H. G. Kazemier, M. L. de Groote, M. H. Ruiters, G. L. Xu and M. G. Rots. 2014. Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res.* **42**(3): 1563-1574.

Cheng, F. F., L. Su and C. Qian. 2016. Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer. *Oncotarget.* **7**: 48832-48841.

Cho, S. W., S. Kim, J. M. Kim and J.-S. Kim. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology.* **31**: 230.

Cho, S. W., S. Kim, Y. Kim, J. Kweon, H. S. Kim and S. Bae. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**: 132-141.

Choi, Y., M. Gehring, L. Johnson, M. Hannon, J. J. Harada, R. B. Goldberg, S. E. Jacobsen and R. L. Fischer. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell.* **110**(1): 33-42.

Choudhury, S. R., Y. Cui, K. Lubecka, B. Stefanska and J. Irudayaraj. 2016. CRISPR-dCas9 mediated TET1 targeting for selective

DNA demethylation at BRCA1 promoter. *Oncotarget.* **7**(29): 46545-46556.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science.*

Cox, D. B. T., J. S. Gootenberg, O. O. Abudayyeh, B. Franklin, M. J. Kellner, J. Joung and F. Zhang. 2017. RNA editing with CRISPR-Cas13. *Science.*

Daboussi, F., S. Leduc, A. Maréchal, G. Dubois, V. Guyot, C. Perez-Michaut, A. Amato, A. Falciatore, A. Juillerat, M. Beurdeley, D. F. Voytas, L. Cavarec and P. Duchateau. 2014. Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nature Communications.* **5**: 3831.

Dagdas, Y. S., J. S. Chen, S. H. Sternberg, J. A. Doudna and A. Yildiz. 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv.* **3**(8).

Dawson, S. J., D. W. Tsui, M. Murtaza, H. Biggs, O. M. Rueda and S. F. Chin. 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med.* **368**: 1199-1209.

De Mattos-Arruda, L. and C. Caldas. 2016. Cell-free circulating tumour DNA as a liquid biopsy in breast cancer. *Mol Oncol.* 10: 464-474.

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

Forbes, S. A., G. Tang, N. Bindal, S. Bamford, E. Dawson and C. Cole. 2010. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* 38: D652-D657.

Forshew, T., M. Murtaza, C. Parkinson, D. Gale, D. W. Tsui and F. Kaper. 2012. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med.* 4: 136ra168.

Freidin, M. B., D. V. Freydina, M. Leung, A. M. Fernandez, A. G. Nicholson and E. Lim. 2015. Circulating tumor DNA outperforms circulating tumor cells for KRAS mutation detection in thoracic malignancies. *Clin Chem.* 61: 1299-1304.

Friedland, A. E., Y. B. Tzur, K. M. Esvelt, M. P. Colaiacovo, G. M. Church and J. A. Calarco. 2013. Heritable genome editing in C.

elegans via a CRISPR-Cas9 system. *Nat Methods.* 10: 741-743.

Fu, W., A. Chaiboonchoe, B. Khraiwesh, D. R. Nelson, D. Al-Khairy, A. Mystikou, A. Alzahmi and K. Salehi-Ashtiani. 2016. Algal Cell Factories: Approaches, Applications, and Potentials. *Mar Drugs.* 14(12).

Gao, H., D. A. Wright, T. Li, Y. Wang, K. Horken, D. P. Weeks, B. Yang and M. H. Spalding. 2014. TALE activation of endogenous genes in Chlamydomonas reinhardtii. *Algal Research.* 5(Supplement C): 52-60.

Garneau, J. E., M.-È. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadán and S. Moineau. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature.* 468: 67.

Gehring, M., J. H. Huh, T. F. Hsieh, J. Penterman, Y. Choi, J. J. Harada, R. B. Goldberg and R. L. Fischer. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell.* 124(3): 495-506.

Harris, E. H. 2001. CHLAMYDOMONAS AS A MODEL ORGANISM.

Annu Rev Plant Physiol Plant Mol Biol. **52**: 363-406.

Heitzer, E., P. Ulz and J. B. Geigl. 2015. Circulating tumor DNA as a liquid biopsy for cancer. *Clin Chem.* **61**: 112-123.

Hilton, I. B., A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy and C. A. Gersbach. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol.* **33**(5): 510-517.

Hsu, P. D., E. S. Lander and F. Zhang. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* **157**(6): 1262-1278.

Ikeda, Y., Y. Kobayashi, A. Yamaguchi, M. Abe and T. Araki. 2007. Molecular basis of late-flowering phenotype caused by dominant epi-alleles of the FWA locus in Arabidopsis. *Plant Cell Physiol.* **48**(2): 205-220.

Jiang, W., A. J. Brueggeman, K. M. Horken, T. M. Plucinak and D. P. Weeks. 2014. Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryot Cell.* **13**(11): 1465-1469.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E.

Charpentier. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science.* 337: 816-821.

Jinek, M., A. East, A. Cheng, S. Lin, E. Ma and J. Doudna. 2013. RNA-programmed genome editing in human cells. *Elife.* 2: e00471.

Jinek, M., 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 and J. A. Doudna. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science.* 343(6176): 1247997.

Kim, D., J. Kim, J. K. Hur, K. W. Been, S. H. Yoon and J. S. Kim. 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol.* 34: 863-868.

Kim, H. and J. S. Kim. 2014. A guide to genome engineering with programmable nucleases. *Nat Rev Genet.* 15: 321-334.

Kim, J. M., D. Kim, S. Kim and J. S. Kim. 2014. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat Commun.* 5: 3157.

Kim, J. S. 2016. Genome editing comes of age. *Nat Protoc.* 11(9):

1573-1578.

Kim, Y. G., J. Cha and S. Chandrasegaran. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A.* **93**(3): 1156-1160.

Kinde, I., J. Wu, N. Papadopoulos, K. W. Kinzler and B. Vogelstein. 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci Usa.* **108**: 9530-9535.

Kinoshita, Y., H. Saze, T. Kinoshita, A. Miura, W. J. Soppe, M. Koornneef and T. Kakutani. 2007. Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J.* **49**(1): 38-45.

Koonin, E. V., K. S. Makarova and F. Zhang. 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology.* **37**(Supplement C): 67-78.

Kruglyak, K. M., E. Lin and F. S. Ong. 2014. Next-generation sequencing in precision oncology: challenges and opportunities. *Expert Rev Mol Diagn.* **14**: 635-637.

Law, J. A. and S. E. Jacobsen. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals.

Nat Rev Genet. **11**(3): 204-220.

Lee, S. H. and S. Bae. 2016. Structural and dynamic views of the CRISPR-Cas system at the single-molecule level. *Bmb Rep.* **49**: 201-207.

Li, M., F. Diehl, D. Dressman, B. Vogelstein and K. W. Kinzler. 2006. BEAMing up for detection and quantification of rare sequence variants. *Nat Methods.* **3**: 95-97.

Liu, X. S., H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young and R. Jaenisch. 2016. Editing DNA Methylation in the Mammalian Genome. *Cell.* **167**(1): 233-247 e217.

Maeder, M. L., J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein, J. F. Costello, M. F. Wilkinson and J. K. Joung. 2013. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol.* **31**(12): 1137-1142.

Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf and E. V. Koonin. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with

eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct.* 1: 7.

Makarova, K. S., Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. Mojica, R. M. Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen and E. V. Koonin. 2015. An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol.* 13(11): 722–736.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church. 2013. RNA–Guided Human Genome Engineering via Cas9. *Science.* 339(6121): 823–826.

Marraffini, L. A. 2015. CRISPR–Cas immunity in prokaryotes. *Nature.* 526(7571): 55–61.

Mok, Y. G., R. Uzawa, J. Lee, G. M. Weiner, B. F. Eichman, R. L. Fischer and J. H. Huh. 2010. Domain structure of the DEMETER 5–methylcytosine DNA glycosylase. *Proceedings of the National Academy of Sciences.* 107(45): 19225–19230.

Moscou, M. J. and A. J. Bogdanove. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science.* 326(5959): 1501.

Narayan, A., N. J. Carriero, S. N. Gettinger, J. Kluytenaar, K. R. Kozak and T. I. Yock. 2012. Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Res.* 72: 3492-3498.

Newman, A. M., S. V. Bratman, J. To, J. F. Wynne, N. C. Eclov and L. A. Modlin. 2014. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med.* 20: 548-554.

Nishimasu, H., F. A. Ran, Patrick D. Hsu, S. Konermann, Soraya I. Shehata, N. Dohmae, R. Ishitani, F. Zhang and O. Nureki. 2014. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell.* 156(5): 935-949.

Nuñez, J. K., P. J. Kranzusch, J. Noeske, A. V. Wright, C. W. Davies and J. A. Doudna. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural &Amp; Molecular Biology.* 21: 528.

Ouyang, M., X. Li, J. Ma, W. Chi, J. Xiao, M. Zou, F. Chen, C. Lu and L. Zhang. 2011. LTD is a protein required for sorting light-harvesting chlorophyll-binding proteins to the chloroplast SRP pathway. *Nat Commun.* 2: 277.

Park, J., K. Lim, J. S. Kim and S. Bae. 2017. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics.* **33**: 286-288.

Park, J. T., N. Johnson, S. Liu, M. Levesque, Y. J. Wang and H. Ho. 2015. Differential in vivo tumorigenicity of diverse KRAS mutations in vertebrate pancreas: a comprehensive survey. *Oncogene.* **34**: 2801-2806.

Park, J. Y., L. J. Kricka and P. Fortina. 2013. Next-generation sequencing in the clinic. *Nat Biotechnol.* **31**: 990-992.

Rudin, N. and J. E. Haber. 1988. Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. *Mol Cell Biol.* **8**(9): 3918-3928.

Schwarzenbach, H., D. S. B. Hoon and K. Pantel. 2011. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer.* **11**: 426-437.

Sizova, I., A. Greiner, M. Awasthi, S. Kateriya and P. Hegemann. 2013. Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases. *Plant J.* **73**(5): 873-882.

Song, C., Y. Liu, R. Fontana, A. Makrigiorgos, H. Mamon and M. H.

Kulke. 2016. Elimination of unaltered DNA in mixed clinical samples via nuclease-assisted minor-allele enrichment. *Nucleic Acids Res.* **44**: e146.

Soppe, W. J., S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef and A. J. Peeters. 2000. The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell.* **6**(4): 791-802.

Sorber, L., K. Zwaenepoel, V. Deschoolmeester, P. E. Y. Van Schil, J. Van Meerbeeck and F. Lardon. 2017. Circulating cell-free nucleic acids and platelets as a liquid biopsy in the provision of personalized therapy for lung cancer patients. *Lung Cancer.* **107**: 100-107.

Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. B. Zhou, L. A. Diaz and K. W. Kinzler. 2013. Cancer genome landscapes. *Science.* **339**: 1546-1558.

Vojta, A., P. Dobrinic, V. Tadic, L. Bockor, P. Korac, B. Julg, M. Klasic and V. Zoldos. 2016. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44**(12): 5615-5628.

Wood, L. D., D. W. Parsons, S. Jones, J. Lin, T. Sjoblom and R. J.

Leary. 2007. The genomic landscapes of human breast and colorectal cancers. *Science.* 318: 1108-1113.

Wu, H. and Y. Zhang. 2014. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell.* 156(1-2): 45-68.

Wu, Y., D. Liang, Y. Wang, M. Bai, W. Tang and S. Bao. 2013. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell.* 13: 659-662.

Xu, X., Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G. L. Xu and R. Hu. 2016. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* 2: 16009.

Xu, X., W. B. Wei, B. Li, F. Gao, Z. Zhang and J. B. Jonas. 2014. Oncogenic GNAQ and GNA11 mutations in uveal melanoma in Chinese. *P Lo S One.* 9: e109699.

Yamano, T., H. Nishimasu, B. Zetsche, H. Hirano, I. M. Slaymaker and Y. Li. 2016. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell.* 165: 949-962.

Yu, J., K. Baek, E. Jin and S. Bae. 2017. DNA-free Genome Editing of Chlamydomonas reinhardtii Using CRISPR and Subsequent Mutant Analysis. *Bio-protocol.* 7, (11) DOI:

10.21769/BioProtoc.2352

Zetsche, B., J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova and P. Essletzbichler. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell.* **163**: 759-771.

Zhang, Y., X. Ge, F. Yang, L. Zhang, J. Zheng and X. Tan. 2014. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep.* **4**: 5405.

요약(국문초록)

크리스퍼-카스 시스템은 세균과 고세균이 가지고 있는 후천성 면역체계로 DNA-RNA 혼성화에 의해서 정확하게 target DNA 혹은 RNA를 찾아 제거하는 시스템이다. 유전자 교정 분야는 크리스퍼-카스가 도입되면서 폭발적인 성장세를 보였으며 이제는 다른 분야에서도 크리스퍼-카스를 이용한 다양한 연구들이 진행되고 있다. 크리스퍼-카스를 다양한 분야에 적용하는 몇가지 연구들을 소개한다.

첫째로 크리스퍼-카스를 이용하여 액체생검 종양조기진단 방법을 개발하였다. CUT-PCR (크리스퍼와 PCR을 이용한 극미량의 타겟 DNA 검출법)이라는 기술은 야생형의 DNA 서열만 절단하는 크리스퍼 엔도뉴클레이즈를 이용하여 극미량의 종양 유래 DNA의 빈도를 높여 검출하는 방법이다. 다양한 크리스퍼 orthologs들을 사용하여 암 연관성 뉴클레오티드 치환 돌연변이들 중 80% 가량의 타겟에 적용이 가능함을 예측할 수 있었고 CUT-PCR 방법을 targeted deep sequencing과 함께 사용하여 다양한 oncogenic 돌연변이를 매우 높은 민감도 (<0.01%)로 검출할 수 있었다. 마지막으로 대장암 환자의 혈액에서 CUT-PCR을 이용해 oncogenic 돌연변이를 검출하여 이 기술이 여러 종류의 암에 대하여 조기 진단에 사용될 수 있음을 보여 주었다.

두번째로는 미세조류 *Chlamydomonas reinhardtii* 에서 외부 DNA를 사용하지 않는 리보뉴클레오프로테인 전달법을 이용하여 *crltd* 돌연변이체를 얻었다. 크리스퍼-카스를 이용하여 얻어진 *crltd* 돌연변이체는 옅은 노란색을 띠고 있으며 세포 내의 클로로필과 PSI 복합체의 양이 야생형에 비해 줄어들어 있었다. 이를 통해 CrLTD는 LHCP 복합체가 엽록체 내의 틸라코이드 멤브레인으로 이동하고

배치되는 데에 관여함을 알 수 있었다. Crltd1 돌연변이체에서 off-target에 의한 mutation을 확인해 보았을 때 off-target에 의한 의도치 않은 변이가 전혀 발생하지 않았다. 본 연구를 통하여 *C. reinhardtii*에서 크리스퍼 리보뉴클레오프로테인 전달법을 최적화하여 보다 쉽게 돌연변이체를 만들 수 있게 되었다.

세번째로 식물의 DNA demethylase인 DME 단백질을 이용하여 targeted demethylation을 유도하는 실험을 진행하였다. DME 단백질의 catalytic domain을 비활성화된 Cas9 단백질과 연결하여 이를 통해 인간세포와 식물에서 특정 지역의 demethylation을 시도하였다. 우선 인간세포에서 *KLF4* 유전자의 CpG island를 타겟팅하는 gRNA와 dCas9-DME를 transfection 하였을 때 유의미한 DNA methyaltion의 변화를 살펴 볼 수 없었다. 식물에서는 *FWA* 유전자를 타겟팅하여 *fwa* epi-mutant를 만들기 위해 FWA 유전자의 반복서열을 타겟팅하는 gRNA와 dCas9-DME를 형질전환식물을 만들어 발현시켰다. T1 형질전환체에서 늦은 개화와 영양생장단계의 axillary meristem을 보이는 개체들의 선발하여 DNA methylation 정도를 분석하였으나 특별한 변화를 보이지 않았고 이 개체들의 표현형은 다음 세대로 전달되지 않았다. 현시점에서 dCas9-DME를 이용한 효율적인 후성유전체 교정도구를 만드는 일은 성공하지 못했으나 추가적인 실험을 통해 개량이 가능할 것이라 생각한다. 후성유전체 교정은 DNA 염기서열에 변화를 주지 않고 형질을 변화시킬 수 있다는 점에서 큰 장점이 있으며 DME 단백질을 이용한 후성유전체 교정은 여전히 많은 가능성을 가지고 있다.

주요어 : 크리스퍼-카스, 유전자 교정, 암 진단, 순환종양DNA, 리보뉴클레오프로테인, DNA 탈메틸화

학 번 : 2007-20351