Sustainable haploidization of human cell lines by CRISPR/Cas-mediated allele-specific deletion of target genes

SOOHYUK CHO

Department of Medical Science

The Graduate School, Yonsei University

Sustainable haploidization of human cell lines by CRISPR/Cas-mediated allele-specific deletion of target genes

Directed by Professor Hyongbum Kim

The Master's Thesis
submitted to the Department of Medical Science
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Master of Medical Science

SOOHYUK CHO

December 2022

This certifies that the Master's Thesis of
SOOHYUK CHO is approved.

Thesis Supervisor : Hyongbum Kim

Thesis Committee Member#1 : Sung-Rae Cho

Thesis Committee Member#2 : Hui-Kwon Kim

The Graduate School
Yonsei University

December 2022

ACKNOWLEDGEMENTS

본 논문을 작성하기까지의 연구 수행 중 도움을 주신 많은 분들께 감사드립니다.

가장 먼저, 유전자 교정 연구에서 선두지위의 연구에 참여할 기회를 주신 저희 김형범 교수님께 감사드리고 싶습니다. 교수님 밑에서 2 년을 거친 연구 경험은 유전자 교정 및 관련 분야를 연구하고자 하는 저의 목표를 더욱 확실히 할 수 있는 계기가 되었습니다. 연구 과정 중 제시해 주신 고견과 다각도의 관점은 추후의 연구 활동에 있어 튼실한 밑거름이 될 수 있도록 하겠습니다.

또한 본 논문의 작성 과정 중 미흡했던 여러 항목들에 대한 의견을 개진해 주시고, 더 의미 있는 연구로 이어질 수 있는 요소에 대해 제언해 주신 조성래 교수님과 김희권 교수님께 감사의 말씀을 드립니다.

유전자 가위를 응용한 유전자의 기능 연구는 저에게 완전히 새로운 분야였습니다. 실험실 구성원들의 수고로운 도움에 힘입어 이 분야에 익숙해지는 데에 많은 시간을 아끼게 되었습니다. 가설에 합당한 실험 방식 및 데이터 접근 방식을 깨우쳐 주신 승호 형, 실험의 기초 방법론에 대한 기반을 잡아 주신 김영광 박사님, 더불어 경험에 입각해 서슴지 않고 다양한 조언을 해주신 여러 실험실 구성원들은 제가 이 분야에 대한 과학적 이해도를 높이고 유연한 관점을 적용하는 데에 큰 도움이 되었습니다. 또한, 분석 결과 해석에 대해 실험적인 측면에서 접근 가능한 여러 가지 의견을 제시해 준 상연이 형과 구상이, 실험 과정에서 놓칠 수 있는 여러 사항들을 공유해준 민영이와 라무를 포함하여 학위과정을 마치는 데에 도움을 주신 모든 실험실 구성원들께 감사드립니다.

제가 여기까지 오기에는, 학부에서부터 받은 수많은 도움이 있었다고 생각합니다. 무엇보다 학사 과정 중에도 연구에 적극적으로 참여할 수 있게 해 주시고, 과학자로서의 진로 설정에 대해 함께 고민해 주신 고혁완 교수님께 감사하다는 말씀드립니다. 또한, 연구자로서의 기본 자세와 과학적 탐구 활동에 필요한 태도를 몸소 보여주신 권오빈 교수님께 깊이 감사드립니다. 마지막으로, 실험에 대한 도전적인 접근 방식과 문제 해결 방식을 깨우쳐 주신 윤홍덕 교수님께 감사의 말씀드립니다. 다양한

분야에서 받은 도움을 바탕으로 보다 폭넓은 연구 경험을 쌓을 수 있었다고 생각합니다. 저는 이제 제가 나아가려고 하는 분야에서 연구를 지속하기 위해 또다른 길을 걷고자 합니다. 많은 분들의 조언에 힘입어, 새로운 길을 잘 헤쳐 나갈 수 있도록 하겠습니다.

끝으로, 제가 걷고자 하는 길을 묵묵히 격려해주시는 부모님과 응원해주는 수성이 모두에게 감사함과 고마움을 느끼며, 사랑하는 가족 모두 지금과 같이 건강하고 꿋꿋하게 앞길을 헤쳐 나갈 수 있기를 기원합니다.

2022 년 10 월

조수혁

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Sustainable haploidization of human cell lines by CRISPR/Cas-mediated allele-specific deletion of target genes

SOOHYUK CHO

Department of Medical Science

The Graduate School, Yonsei University

(Directed by Professor Hyongbum Kim)

CRISPR/Cas-system is one of the most powerful tools that enable biological functional screening via precise genome editing at the intended target regions. Up to date, base editors (BEs) and prime editors (PEs) are the most frequently used genome editing tools in inducing specific mutations at the target loci to demonstrate its relevance to the resulting phenotypes within host, both *in vitro* and *in vivo*. In line with this, recent development of haploid cell lines, including KBM7 and HAP1, has also enabled efficient genome screening by preventing unedited alleles from masking the phenotype changes. However, such cell lines showed unstable haploidy, being enriched in diploids during their long-term culture. Therefore, instead of using near haploid cell lines, introducing haplotype at gene-of-interest in diploids is expected to generate more sustainable haploidy. Targeting single nucleotide polymorphisms (SNPs)-enriched region would provide robust specificity to induce haplotype at the gene-of-interest within human diploid. By single cell cloning the human cell lines that exhibit CRISPR/Cas-mediated large deletions, diploids with sustainable haploidy at the target

genes are expected to be generated.

Here, we induced large deletion in BRCA2 by CRISPR/Cas9-mediated DNA double strand break, precisely targeting the DNA binding domain of BRCA2. The real-time quantitative PCR (RT-qPCR) results demonstrate the robust efficiency of large deletion in the target loci, resulting in gene-specific haploidy of BRCA2, which could be the ideal target for efficient functional screening. Relative expression levels of large deletion-induced BRCA2 in single cell samples show one sample with induced haploidy of BRCA2, which was identified by RT-qPCR amplification of each single cell. Although sustainability of its haploidy should be continuously monitored during a long-term culture, subsequent deep sequencing of haploidized BRCA2 loci would further validate its genotype in a sequence-specific manner.

Keywords: BRCA2, CRISPR/Cas9, large deletion, haploidization, dual guide RNA

I. INTRODUCTION

The development of targeted genome editing via CRISPR/Cas9 has expanded the scope of functional screening of multiple gene sets, which had been practiced in a limited measure since the Human Genome Project[1, 2]. Up to date, genome editing is no longer tied to gene knock out or knock down as a whole, and it expands to a single base level in which a target base is substituted with the desired base at the precise position[3]. Base editors (BEs), the active domain of which is a deaminase fused to catalytically impaired Cas protein[3], are the examples that serve this purpose by altering the original base into induced single base polymorphisms (SNPs). Prime editors (PEs), the most recently developed genome editing tools, modify target genome without double-stranded DNA breaks, and therefore, do not require additional donor DNA as a template[4]. Instead, PEs utilize RNA templates integrated with the guide RNAs to enable nucleotide substitutions or structural variations, including insertions and deletions, within the genome. Their advantages over base editors, including reduced off-target effects and increased options for edit types and windows in genome editing, have led to active application of prime editors in recent research[5,6].

The advent of haploid cell lines in the field of genome screening has also promoted the application of CRISPR/Cas-system to studying gene function. The monoallelic gene expression of the haploids is known to prevent unedited alleles from mitigating the changes in cell phenotypes, especially in heterozygotes[7]. Given that a single allele is sufficient to exhibit the resulting phenotypes of mutations for most of the genes, haploids have emerged as the promising option for efficient genome screening. Near haploids isolated from a patient with chronic myelogenous leukemia, including KBM7 and HAP1, however, showed unstable haploidy because they became enriched in diploids during long-term cultures. The observed diploidization could be partly due to a proliferative advantage of diploid cells over haploid cells, but further research is required.

The unstable karyotypes of haploids have been classified as a common phenomenon during haploid cultures. For example, HAP1 cells rapidly decreased in the percentage of haploids[8], which was also observed in other haploid human cells[9]. Up to date, the uncoordinated centrosome cycle in haploids has been identified as a possible reason for the loss of HAP1 haploidy[10]. In addition, the activation of p53-dependent response in haploids has been identified as the cause for their reduced viability during culture[8]. Therefore, fundamental methods to prevent diploidization are necessary for CRISPR/Cas-mediated genome editing within haploid cell lines. Instead of using near haploid cell lines, selectively introducing haploidy to the target genes of diploid cell lines could be an alternative option to maintain the haploidy. In previous studies, precise allele-specific cleavage of target genes has been observed, using guide RNAs with protospacer adjacent motif (PAM) sites harboring the allele-specific single nucleotide polymorphisms (SNPs)[11, 12]. Therefore, large deletion of certain alleles would be possible if additional guide RNA cleaves the other end of the target gene in a pairwise manner.

II. MATERIALS AND METHODS

1. Cell culture and transfection conditions

HEK293T cells were cultured and passaged in DMEM supplemented with 10% (vol/vol) fetal bovine serum at 37 °C in the presence of 5% CO2. For transfection, when cells reach 60%–70% confluency, SpCas9 and sgRNA plasmid were transfected using Lipofectamine 2000 (Thermo Fisher Scientific), according to the manufacturer's protocol. The cells were incubated overnight, after which they were maintained in 2 μg ml−1 of puromycin for the next 5 d to remove untransfected cells. 5 days after transfection, half of cell pellets were homogenized, and genomic DNA of the cells were extracted using lysis buffer and proteinase K. To genotype the samples, the BRCA2, HPRT1 regions of interest were amplified, and PCR products were sequenced by Sanger sequencing.

2. Detection and validation of SNP sites

All the single nucleotide polymorphisms (SNPs) were determined by integrative genome viewer (IGV), a SNP calling software provided on-line, referring to hg38 reference genome data. For genomic DNA information of HEK293T, single-cell data was obtained from a previous study on genome dynamics of human embryonic kidney cells[13]. After choosing SNP-enriched regions in BRCA2, lab cultured HEK293T batch was sequenced for further validation.

3. Dual guide RNA design

All guide RNAs targeting BRCA2 were designed based on the scoring method of DeepSpCas9, a deep learning-based model from previous study[14], for predicting the efficiency of the guide RNAs. Guide RNAs that scored above 50 were selected for future experiments. For pan-allelic guide RNAs, the target loci were placed in the upstream and downstream of exon 17 and exon 18, respectively, while SNP-specific

guide RNA was placed in the downstream of exon 18 harboring hetero SNP that was validated through individual sequencing.

4. Single cell culturing conditions

For samples treated with pan-allelic deletion method, single cells were isolated into 96-well plates by using fluorescence-activated cell sorting (FACS) Aria2 (BD Biosciences). When cells reached 60% confluency during culture, they were transferred to 24-well plates and 6-well plates, subsequently. After reaching 80% confluency, cell pellets were homogenized, and genomic DNA of each clone was extracted using lysis buffer and proteinase K.

5. RT-qPCR for identifying copy number variations

RT-qPCR was performed using an ABI 7700 Sequence Detection System (PE Applied-Biosystems) in the presence of SYBR-green. The optimization of the real-time PCR reaction was performed according to the manufacturer's instructions (PE Applied-Biosystems, User Bulletin 2 applied to the SYBR-Green I core reagent protocol) but scaled down to 20 μl per reaction, with 25ng of genomic DNA as a template. The PCR conditions were standard (SYBR-Green I core reagent protocol) and all reagents were provided in the SYBR-Green I core reagent kit, including AmpliTaq-GOLD polymerase (PE Applied-Biosystems). The primer sets for RT-qPCR were designed and selected through PrimerQuest Tool, according to Integrated DNA Technologies (IDT)'s instruction (Table 2). Relative gene expression was calculated based on relative cycle threshold (CT) values of exon 17, comparing to the CT values of their neighboring exon 16 as a reference.

6. Large amplicon deep sequencing

The proportion of allele-specific large deletion was calculated based on the read depth of long template PCR, size of which is up to 15kb including the target gene. The long template PCR product was sheared using a Covaris LE220 to an average size of 300 bp and purified using Ampure XPRI beads according to the manufacturer's

protocol. Randomly sheared DNA was end-repaired, A-tailed, and ligated to adapters containing an 8nt unique molecular identifier. Samples were amplified with two rounds of nested PCR with primers that complement the oligo tag. The sample libraries were sequenced on Illumina MiniSeq platform.

III. RESULTS

1. IGV alignment for SNP enriched sites in BRCA2

Based on the single nucleotide polymorphisms (SNPs) database of human embryonic kidney cells from a previous study[13], IGV software identified hetero SNP-enriched loci within BRCA2′s downstream loci of exon 18, which was further validated by Sanger sequencing (Fig. 1A). For allele-specific deletion of BRCA2, guide RNA 2 candidate was designed to target the hetero allele without SNP (Fig. 1B). Guide RNA 1 candidates targeted upstream loci of exon 17 in all alleles, in order to induce large deletion when applied with guide RNA 2 (Fig. 1B). The pool of guide RNA candidates was further applied in a pairwise manner, targeting both the upstream and downstream of exon 17 and 18. As each of the guide RNA candidates were filtered by DeepSpCas9 score (Table 1), the overall estimate of Cas9 efficiency was aimed for 50~60%.

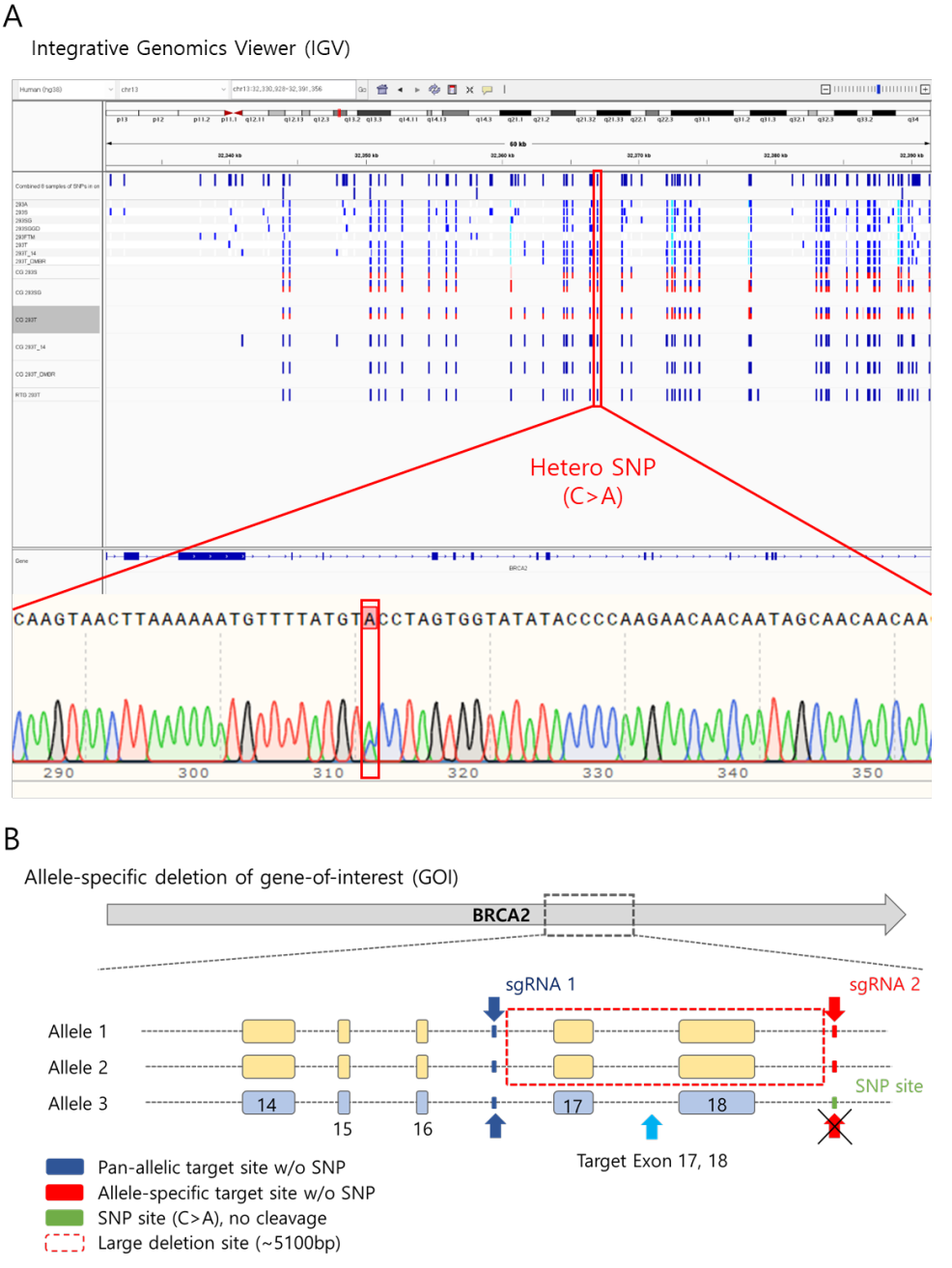

Figure 1. Identification of SNP-enriched sites in BRCA2. (A) IGV software showing the entire SNP sites from human embryonic kidney cell database. (B) Schematics of designing dual guide RNAs, consisting of pan-allelic guide RNA that targets all alleles, and allele-specific guide RNA that targets the alleles without SNP.

Table 1. Guide RNA sequence information. DeepSpCas9 score was calculated based on the on-line webtool from previous study[14].

Guide RNA	Sequence	DeepSpCas9 score
sgRNA 1-1	GGATATGGAAGTCTATACAC	76.36
sgRNA 1-2	CTGTAGTCTAAGTGGAAATG	61.62
sgRNA 2-1	TCTTGGGGTATATACCACTA	67.98
sgRNA 2-2	GTATTAGTAGAGGGAAGATG	65.33

2. Validation of guide RNA activities on target sites

In order to validate whether the designed guide RNAs function in human cell lines, individual guide RNAs were transfected to HEK293T cells with Cas9 and sequenced in Illumina next generation sequencing (NGS) platform. Unlike the expectation, however, each guide RNA showed cleavage activity below 30%, which was half the efficiency of the initial goal (Fig. 2A). This put more strength on the other approach, where all the alleles are targeted by dual guide RNAs that are both pan-allelic, leaving one allele after Cas9-mediated cleavage in a stochastic manner. Therefore, further experiments treated HEK293T cells with either allele-specific or pan-allelic guide RNAs (Fig. 2B).

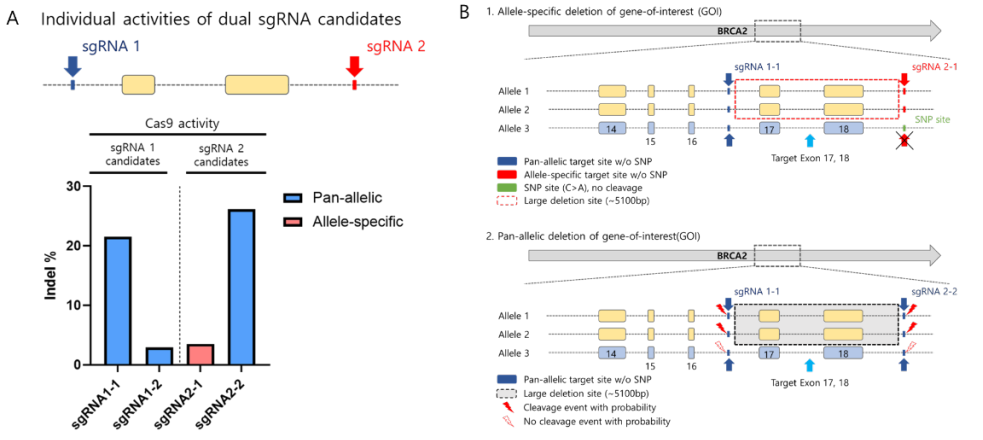

Figure 2. Individual cleavage activities of guide RNAs. (A) Cas9 activities for individual guide RNAs in HEK293T cell lines. (B) Different approaches of large deletion in BRCA2 by allele-specific and pan-allelic dual guide RNAs.

3. Detection of large deletion in BRCA2

After treating HEK293T cells with Cas9 and dual guide RNAs for 5 days, the presence of the large deletion was confirmed by gene amplification of BRCA2 by out-out primer-based PCR amplification (Fig. 3A and Table 2). The single linear band provides evidence of large deletion at the intended loci, by its unique size and sequence data (Fig. 3B). This result shows that dual guide RNAs can make precise large deletions by cutting both ends of the target gene yet limited by its efficiency for certain guide RNA candidates (Fig. 2A). This promoted us to further validate whether the allele frequency of BRCA2 is measurable by RT-qPCR.

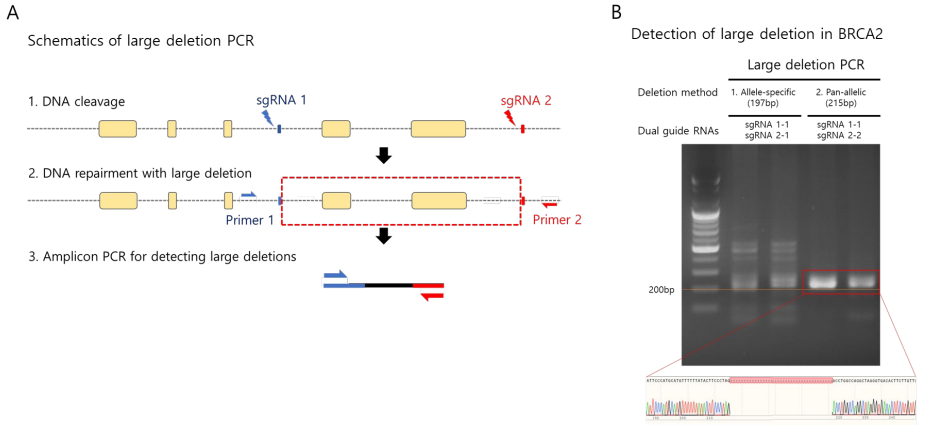

Figure 3. Large deletion detection by PCR amplification. (A) Gene amplification schematics of out-out PCR amplification. (B) The intended large deletion amplification confirmed by its sequence.

Table 2. Primer set for large deletion detection. The primers were designed to anneal to the upstream and downstream of expected large deletion site.

Primer	Sequence	Position
BRCA2_del_F	GGAGGGAAATGGAGAGTTTGG	Intron 16
BRCA2_del_R	ATGTGATCACAGTCGATTAGG	Intron 18

4. Mean allele frequency of target exon in pooled cell batch

Since the cost for deep sequencing each of the single cells in Illumina Miniseq settings is demanding, it was necessary to first identify the cells that show decreased allele frequency results from RT-qPCR method. For proof-of-concept experiment, the mean allele frequency of the pooled cell batch of each experiment was obtained by RT-qPCR (Fig. 4 and Table 3). The result showed up to 43% decrease in allele frequency of the pan-allelic deletion samples (Fig. 4). However, RT-qPCR data shows the allele frequency in an intensity-based manner, not a sequence-specific manner. This made us to further navigate deep sequencing data of the pooled cell batch of HEK293T cells.

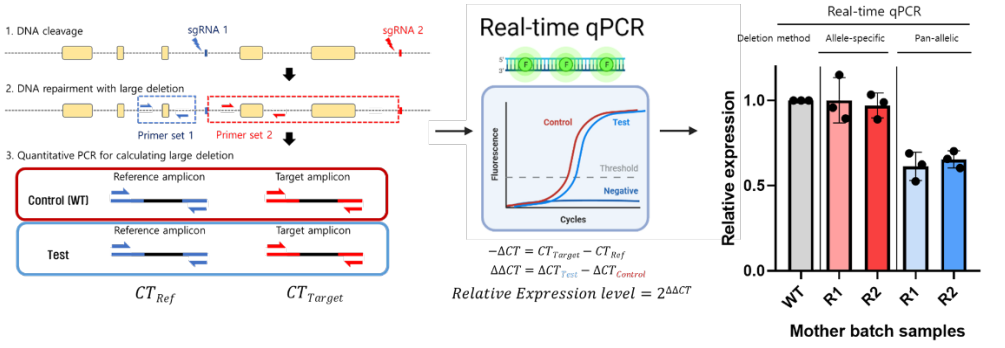

Figure 4. Relative expression of target exon by RT-qPCR. Reference and target amplicons were separately quantified by RT-qPCR. Mean values of RT-qPCR represent relative expression of BRCA2 in control samples, allelic deletion samples, and pan-allelic deletion samples in technical replicates (n=3). R1 and R2 indicate biological replicate 1 and 2, respectively.

Table 3. Primer sets for RT-qPCR amplification. The primers for target site and reference site were designed separately for precise comparison.

Primer	Sequence	Target exon
BRCA2_qpcr_F1	ACGTATGGCGTTTCTAAACATTG	Exon 16
BRCA2_qpcr_R1	TCCAGTCCATAAACTTTCCTTACC	Exon 16
BRCA2_qpcr_F2	TATGGAAACTGGCAGCTATGG	Exon 17
BRCA2_qpcr_R2	AGTTGAAGAAGCACCCTTTCT	Exon 17

5. Quantification of large deletion in BRCA2 by deep sequencing

For deep sequencing of exon 17 and exon 18 loci, where large deletion was partially identified by RT-qPCR, the 15kb-long region including the target exons was amplified by long-range PCR (Fig. 5A and Table 4). After mechanical shearing of PCR product, deep sequencing of DNA fragments was performed through Illumina Miniseq instrument. The proportion of large deletion was calculated by quantitative comparison between the reads with intended deletion and reads with intact DNA sequence spanning the deletion region. The results show that pan-allelic large deletion method was able to introduce up to 36.7% of the intended large deletion to the target site (Fig. 5B). In addition, the efficiency of large deletion was calculated by the mean coverage loss of the target exon 17 and 18, compared to the neighboring regions of the deletion site. To achieve this, bwa algorithm was used for reference genome alignment and samtools coverage function was used for analyzing the read coverage per base position. For the sample treated with pan-allelic guide RNAs, the mean coverage of deletion site was shown to be only 85% of the neighboring regions′ mean coverage (Fig. 5C). This contrasts with the untreated control sample, which shows the mean coverage of deletion site up to 97%, compared to that of the neighboring regions (Fig. 5C). This result shows that pan-allelic deletion method induces large deletion in the target exon 17 and 18 of BRCA2. Therefore, we decided to proceed single cell culture of the HEK293T cell batch with pan-allelic deletion, which we will refer to batch A sample for convenience from now on (Fig. 6A).

Table 4. Primer set for long-range PCR of BRCA2. The primers were designed to amplify 15kb-long BRCA2 region including exon 17 and 18.

Primer	Sequence	Position
BRCA2_ngs_F	CTGAGGTATTAGGCGGAGTAGAGAGTTCAG	Intron 13
BRCA2_ngs_R	AGGAAAAGGTCTAGGGTCAGGAAAGAATCC	Exon 19

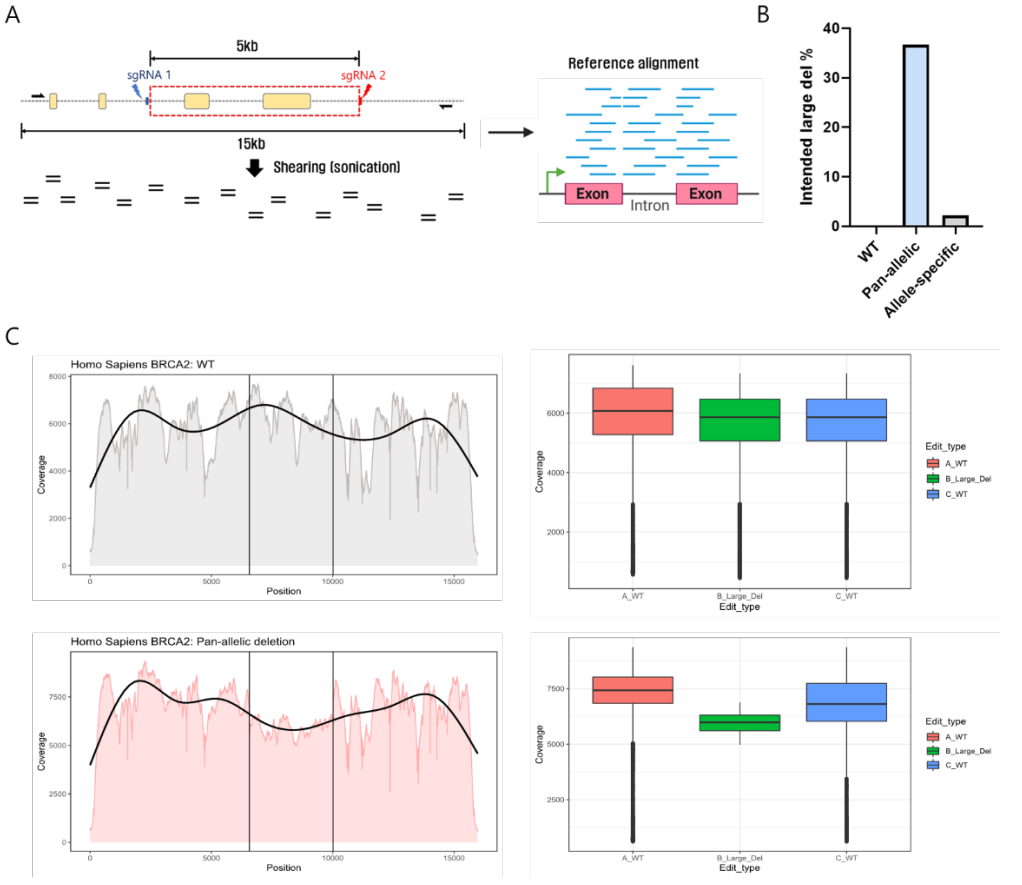

Figure 5. Mean coverage calculation of BRCA2. (A) Long-range PCR amplified 15kb-long region of BRCA2, including target exon 17 and 18. (B) Quantitative comparison between reads with intended large deletion and reads without deletion shows large deletion identified in pan-allelic deletion sample. (C) The gray area represents the coverage per base position in untreated control sample, while red area represents that of pan-allelic deletion sample. Boxplot shows the coverage distribution classified by the deletion site named B_Large_Del, and its neighboring sites, A_WT and C_WT, either left or right to the deletion site, respectively. Reference genome alignment was done by bwa, and the coverage calculation was done by samtools.

6. Single cell culture and validation of large deletion in BRCA2

The batch A samples were sorted into single cells in 96 well plates and were cultured for clonal expansion for the following four weeks (Fig. 6A). After reaching individual confluency of 60%, the clones were transferred to 24-well plates and 6-well plates at biweekly and weekly intervals, respectively. When clones reached 80% confluency in 6-well plates, the samples were individually harvested and stocked for further experiments (Fig. 6B). In parallel with the batch A, we also generated batch B, which were derived from the batch A with extended treatment of pan-allelic guide RNAs (Fig. 6A). Overall, we were able to harvest 66 single cells from batch A and 32 single cells from batch B. To validate the copy numbers of BRCA2 in each sample, we conducted RT-qPCR and identified samples with reduced copy number of BRCA2 based on their relative expression levels. Unexpectedly, the RT-qPCR results of 10 samples from batch A showed no decrease in copy number of BRCA2 (Fig. 7A). This result may be due to insufficient time for dual guide RNAs to induce efficient large deletions in human cells. In contrast, the RT-qPCR results of batch B samples showed decreased copy number of BRCA2 in several samples (Fig. 7B). Among those samples, B#17 showed one third of relative expression in average, compared to that of the untreated control. This indicates that B#17 would likely harbor haploidy in target exon 17 and 18 of BRCA2.

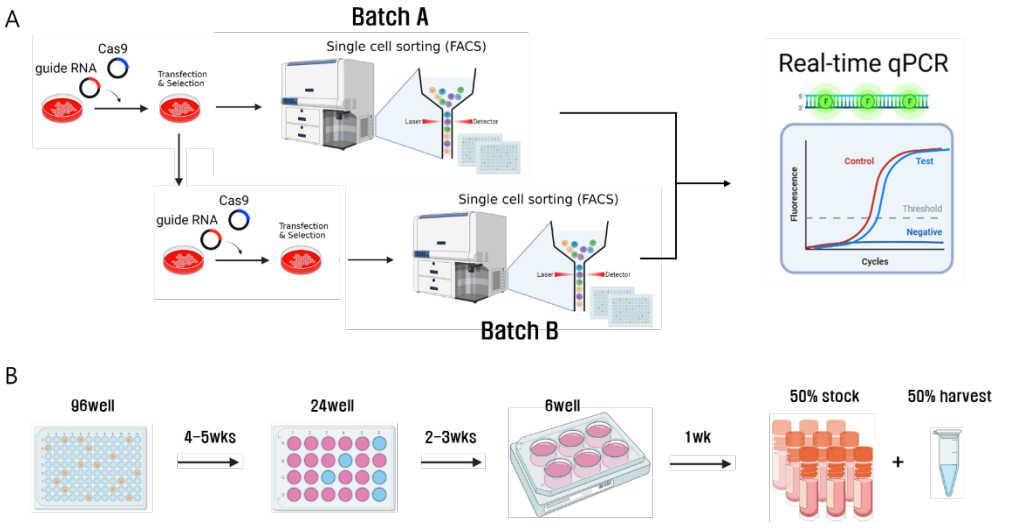

Figure 6. Single cell sorting process of batch A and batch B. (A) Batch A was treated with pan-allelic guide RNAs for 5 days and batch B was treated for additional 5 days. (B) Single cells went through clonal expansion until each sample reached 80% confluency in 6-well plates before harvest.

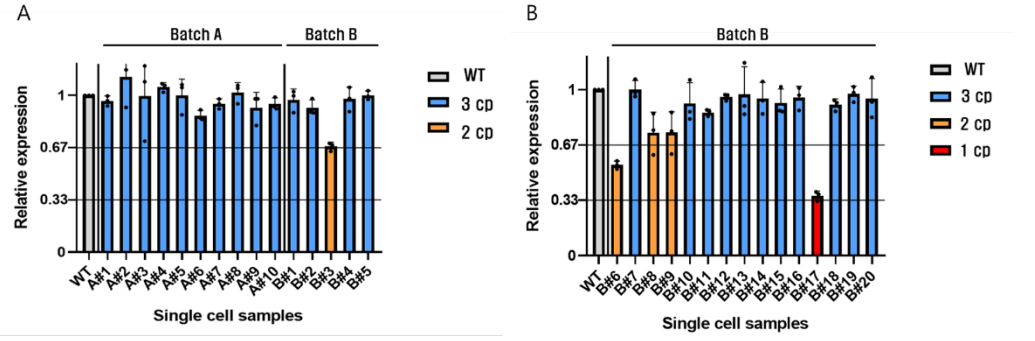

Figure 7. Relative expression of target exon in single cell samples. (A), (B) In 96 well plate, one untreated control and 15 samples were tested in triplicate (n=3). Mean values of RT-qPCR represent relative expression of BRCA2 in untreated controls and single cells from batch A and B. The relative expression levels were normalized to the control, which equals to copy number 3 in BRCA2. The copy numbers of each sample were labeled with separate colors. Blue, orange, and red indicate 3 copies, 2 copies, and 1 copy of BRCA2 in each sample, respectively.

IV. DISCUSSION

There have been several attempts to induce large deletions in gene of interest in previous studies[15, 16]. However, there is yet to be a gold standard for inducing large deletions in a quantitative manner, which could be applied to generate haploidy of target genes in human cell lines for efficient genome screening. This makes the attempt to induce 5kb-long deletion in BRCA2 valuable. The large deletion efficiencies in a pooled cell batch shows the validity of gene-specific large deletions by dual guide RNAs. Although allele-specific guide RNA did not show robust activity in our initial attempt, we show that pan-allelic deletion could be an alternative approach to achieve large deletion in target genes.

RT-qPCR results of single cells from batch A and B indicates that 5 days of treatment with dual guide RNAs could be insufficient to show large deletions in the intended region, and we therefore recommend at least 10 days of treatment. Interestingly, RT-qPCR results from single cell samples had discrete distribution of relative expression level, ranging from one-fold to three-fold, and this indicates that RT-qPCR is capable of predicting copy number variation of target genes in a DNA level.

Although we observed a single cell sample with mono-allelic expression of BRCA2 by RT-qPCR, further deep sequencing of the target region is required to validate its haploidy in a sequence-specific manner. Moreover, continuous checkup for its haploidy by both RT-qPCR and deep sequencing should be followed to demonstrate the sustainability of its haploidy in BRCA2. In concordance, the deletion size of target gene also needs to be scaled up more than 5kb, which covers only a few exons within a gene. With such size limitation, the deletion sites will be tied to the exons that are known to be essential in gene function. For extended usage of gene-specific haploid generation, deletion size should be even expanded above 20kb, which could enable knocking out accessory alleles at a gene-level, encompassing all the exons required for functional gene expression.

For future application of the generated haploid cell line, functional screening of disruption or mutagenesis in exon 17 and 18 of BRCA2 could be the promising goals. Since functional screening of gene perturbation usually requires haplo-insufficiency of the target genes when the editing efficiency on the targets are limited. By introducing haploidy to such regions, therefore, will solve such masking effect of unedited allele over edited allele, when performing the functional screening experiments.

V. CONCLUSION

The interplay between Cas9 and dual guide RNAs shows the validity of dual guide RNA-mediated large deletion in the target loci. Considering the limited efficiency of individual guide RNAs, there is still a room for improvement in haploidization efficiency of human cell lines, by finding better targets and cell lines that harbor higher efficiency in terms of Cas9 activity.

As newer versions of Cas9 are being developed continuously[16, 17], different approaches could take place to make the haploid cell lines, instead of using wildtype Cas9 to simply induce double strand break into both ends of target genes. For example, paired prime editing is one of the promising tools for inducing up to 10kb-long deletion, which covers more exons than 5kb, in the targeted genome loci[16]. Yet, those methods also require further improvements, in terms of the efficiency of large insertion or deletion, which is down to 1%, not efficient enough to induce haploidization of target genes. Therefore, concurrent development of Cas9 variants would make gene-specific haploidization even more possible in the following future.

REFERENCES

1. Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun 2018. 9(1): p. 1-15.
2. Shalem O, Sanjana NE, Shi X, Scott DA, Mikkelsen TS, Heckl D et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014. 343(6166): p. 84-87.
3. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016. 533(7603): p. 420-424.
4. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 2019. 576(7785): p. 149-157.
5. Zhi S, Chen Y, Wu G, Wen J, Wu J, Liu Q et al. Dual-AAV delivering split prime editor system for in vivo genome editing. Mol Therapy 2021: p. 283-294.
6. Park SJ, Jeong TY, Shin SK, Yoon DE, Lim SY, Kim SP et al. Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. BMC 2021. 22(1): p. 1-11.
7. Li Y, Shuai L. A versatile genetic tool: haploid cells. Stem Cell Res Ther 8 2017 197 (2017).
8. Olbrich T, Mayor-Ruiz C, Vega-Sendino M, Gomez C, Ortega S, Ruiz S et al. A p53-dependent response limits the viability of mammalian haploid cells. PNAS 2017. 114(35): 9367-9372.
9. Yilmaz A, Peretz M, Sagi I. Haploid Human Embryonic Stem Cells: Half the Genome, Double the Value. Cell stem cell vol. 19,5 (2016): 569-572. doi:10.1016/j.stem.2016.10.009
10. Yaguchi K, Yamamoto T, Matsui R, Tsukada Y, Shibanuma A, Kamimura K et al. Uncoordinated centrosome cycle underlies the instability of non-diploid

somatic cells in mammals. The Journal of cell biology vol. 217,7 (2018): 2463-2483. doi:10.1083/jcb.201701151

11. Erwood S, Bily T MI, Lequyer J, Yan J, Gulati N, Brewer RA et al. Saturation variant interpretation using CRISPR prime editing. Nat Biotechnol 40, 885–895 (2022).
12. Rabai A, Reisser L, Reina-San-Martin B, Mamchaoui K, Cowling BS, Nicot AS et al. Allele-Specific CRISPR/Cas9 Correction of a Heterozygous DNM2 Mutation Rescues Centronuclear Myopathy Cell Phenotypes. Molecular therapy. Nucleic acids vol. 16 (2019): 246-256. doi:10.1016/j.omtn.2019.02.019
13. Lin, YC, Boone M, Meuris L, Lemmens I, Roy NV, Soete A et al. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nat Commun 5, 4767 (2014).
14. Kim H, Kim Y, Lee S, Min S, Bae J, Choi J et al. SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. Science advances vol. 5,11 eaax9249. 6 Nov. 2019, doi:10.1126/sciadv.aax9249
15. Anzalone AV, Gao XD, Podracky CJ, Nelson AT, Koblan LW, Raguram A et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat Biotechnol 40, 731–740 (2022).
16. Choi J, Chen W, Suiter CC, Lee C, Chardon FM, Yang W et al. Precise genomic deletions using paired prime editing. Nat Biotechnol 40, 218–226 (2022). https://doi.org/10.1038/s41587-021-01025-z
17. Peter JC, Hussman JA, Yan J, Knipping F, Ravisankar P, Chen PF et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell, Volume 184, Issue 22: p5635-5652.e29 (2021).

IIV. ABSTRACT (IN KOREAN)

CRISPR/Cas 시스템의 allele 특이적 유전자 deletion을 이용한
지속가능한 하플로타입 인간세포주 제작

<지도교수 김형범>

연세대학교 대학원 의과학과

조 수 혁

CRISPR/Cas 시스템은 표적 영역에서의 정밀한 유전자 편집을 통해 유전자의 생물학적, 기능적 스크리닝을 가능하게 하는 가장 강력한 도구이다. 최근에는, 염기 교정자 및 프라임 교정자라는 유전자 가위가 생체 내에서 유전형과 표현형과의 관련성을 입증하기 위해 표적 위치에서 특정 염기를 삽입 또는 삭제하도록 응용되고 있다. 이와 맞물려 최근 KBM7과 HAP1을 비롯한 반수체 (n) 세포주가 개발됨에 따라, 그동안 교정되지 않은 대립유전자가 표현형 변화를 가리는 현상을 방지하며 효율적인 표현형 스크리닝이 가능해졌다. 그러나 이러한 반수체 세포주의 배양 과정에서 이배체가 빠르게 확산되는 것으로 미루어, 반수체 상태가 불안정한 것으로 드러났다. 따라서, 이러한 반수체 세포주를 사용하는 대신, 이배체 세포주 내에서 특정 유전자에 한해 반수체를 도입하면 안정적으로 반수성을 유지할 수 있을 것으로 기대된다. 단일 뉴클레오타이드 다형성(SNP)이 풍부한 영역을 대상으로 유전자 가위를 표적 것은 인간 이배체 내에서 관심 유전자에서 하플로타입을 유도하는 강력한 특이성을 제공할 것이다. 유전자 결실을 나타내는 인간 세포주를 단세포 수준에서 클로닝함으로써, 표적 유전자에서 지속 가능한 반수체를 갖는 이배체가 생성될 것으로 예상된다.

본 연구에서는 CRISPR/Cas9 매개 DNA 이중 가닥 파괴에 의해 표적 유전자에서 대규모 유전자 결실을 유도했는데, 그 중에서도 BRCA2의 DNA 결합 도메인을 대상으로 한다. 단일 세포 샘플들의 실시간 정량적 PCR 결과는 표현형 스크리닝의 주요 대상이 될 수 있는 단일 대립 유전자를 제외하고, 해당 위치에서 대규모 유전자 결실을 일으킬 수 있음을 보여준다. 본 연구에서 확보한 단일 세포주의 BRCA2 유전자의 반수성 유지에 대한 검증을 위해서는, 해당 세포의 장기간 배양 및 수차례 검증 실험이 필요할 것으로 예상된다.

핵심되는 말: CRISPR/Cas9, 유전자 결실, 하플로타입, 이중 sgRNA