scientific reports

OPEN

The efficacy of CRISPR-mediated cytosine base editing with the RPS5a promoter in *Arabidopsis thaliana*

Minkyung Choi[1,4], Jae-Young Yun [ID][1,2,4], Jun-Hyuk Kim[3], Jin-Soo Kim [ID][1✉] & Sang-Tae Kim [ID][1,3✉]

CRISPR/Cas9-mediated genome editing is an important and versatile technology in modern biological research. Recent advancements include base-editing CRISPR tools that enable targeted nucleotide substitutions using a fusion protein comprising a nickase variant of Cas9 and a base deaminase. Improvements in base editing efficiencies and inheritable of edited loci need to be made to make CRISPR a viable system in plants. Here, we report efficiency of cytosine base editors (CBEs) in *Arabidopsis thaliana* by applying the strong endogenous RPS5a promoter to drive the expression of nickase Cas9 and either rAPOBEC1 from rat (BE3) or the PmCDA1 activation-induced cytidine deaminase from sea lamprey (AIDv2). Compared with the strong heterologous CaMV35S promoter of viral origin, the RPS5a promoter improved CBE efficiency by 32% points with the number of T$_1$ plants showing over 50% conversion ratio when the *LFY* gene was targeted. CBE induced nonsense mutations in *LFY* via C-to-T conversion, which resulted in loss-of-function *lfy* phenotypes; defects in *LFY* function were associated with the targeted base substitutions. Our data suggest that optimal promoter choice for CBE expression may affect base-editing efficiencies in plants. The results provide a strategy to optimize low-efficiency base editors and demonstrate their applicability for functional assays and trait development in crop research.

Clustered regularly interspaced palindromic repeat (CRISPR) and CRISPR-associated protein 9 (Cas9)-mediated genome editing technologies provide a novel platform for biological research by enabling precise and program-mable modification of genetic elements[1–3]. CRISPR/Cas9-mediated genome engineering functions by introducing a DNA double-strand break 3 base pairs (bp) upstream of the protospacer adjacent motif (PAM), to which Cas9 endonuclease is recruited via a single stranded guide RNA (gRNA). The double-strand break is often repaired by the error-prone non-homologous end joining (NHEJ) process; NHEJ can generate a small insertion or dele-tion (indel) during repair. CRISPR/Cas9-mediated indels often introduce an early stop codon by a frameshift, rendering the target gene non-functional. Although CRISPR/Cas9 is a powerful tool for knocking out genes, typically by generating loss-of-function alleles, introducing specific amino acid changes to a target gene can be difficult due to technical limitations; this has limited functional genomics research in which specific mutations are desired to address sub- or neo-functionalization of a given gene product.

Base editors that can function in eukaryotic cellular systems with various base deaminases fused to CRISPR/Cas9 have recently been developed, thus broadening the gene modification toolbox for functional genomics[4–6]. Cytosine base editors (conversion of C/G to T/A; CBEs, with partial conversion of C/G to G/C unexpected) can introduce base substitutions that lead to amino acid changes or splicing modifications. Two CBE systems, base editor 3 (BE3) and activation-induced cytidine deaminase version 2 (AIDv2) contain the adapted cytidine deaminases rAPOBEC1 from rat and PmCDA1 from sea lamprey, respectively[5,7]. Generally, the base editing target window of BE3 is narrower than that of AIDv2; BE3 has 4–8 positions on the 20 bp target sequences, while AIDv2 has 2–10 positions when counting the canonical PAM sequence (NGG) as at positions 21–23.

CBEs have been applied in various plant species, ranging from monocots such as *Oryza sativa*, to the model plant eudicot *Arabidopsis thaliana* (hereafter, *Arabidopsis)*. However, the overall success rate for base editing

[1]Center for Genome Engineering, Institute for Basic Science, Daejeon 34126, Republic of Korea. [2]Institute of Green Bioscience & Technology, Seoul National University, Pyeongchang, Republic of Korea. [3]Department of Medical and Biological Sciences, The Catholic University of Korea, Bucheon 14662, Republic of Korea. [4]These authors contributed equally: Minkyung Choi and Jae-Young Yun. ✉email: jskim01@snu.ac.kr; stkim@catholic.ac.kr

using CBEs remains relatively low compared to CRISPR/Cas9[8–14]. Parameters affecting genome editing efficiencies in plant cells have been determined using stable transgenic *Arabidopsis* lines, into which the conventional CRISPR/Cas9 cassette was introduced via *Agrobacterium*-mediated transformation[15–17]. Promoter choice is an important factor that affects the indel-inducing efficiencies of Cas9 in plant cells. Both CaMV35S (hereafter 35S) and ubiquitin promoters, which are widely used in plant transformation practices, have resulted in relatively high Cas9 expression in plant cells with moderate success rates. However, recent evidence suggests that alternative promoter systems may be more suitable for CRISPR/Cas9 expression and for generating stable transgenic lines. For example, the ribosomal protein subunit 5a (RPS5a) promoter of *Arabidopsis* has recently been shown to achieve higher efficiencies of adenine base editing and more successful transmission of edited alleles into the next generation in *Arabidopsis* compared to canonical promoters[18]. Similar improvements were also observed in regular CRISPR/Cas9 mediated mutagenesis using the RPS5a promoter[15,16]. However, the effectiveness of the RPS5a promoter for CBEs has not yet been addressed.

In this study, we aimed to test the efficacy of the RPS5a promoter in CBEs and to assess how promoter choice can influence the efficiencies of two CBE systems: BE3 and AIDv2. We selected the *LEAFY* (*LFY*) gene, a master transcription factor for flowering, as a CBE target to address the efficiency of CBE-mediated base editing in *Arabidopsis*. Disruption of *LFY* function can be easily identified by characteristic knock-out phenotypes (*lfy*), which include the loss of floral organs such as petals and stamens[19–21]. We measured the efficiencies of four CBE vector constructs with combinations of two different promoters, RPS5a and 35S, to drive the expression of the two CBE variants; the CBE systems were designed to introduce nonsense point mutations of cytosine (C) to thymine (T) in two target sites within the *LFY* gene. Clear *lfy* mutant phenotypes were observed in both the first (T_1) and second (T_2) generations of transgenic lines. A next-generation sequencing (NGS)-assisted large-scale survey of the C-to-T conversion efficiencies in 440 transgenic T_1 individuals enabled us to assess the base editing patterns of the two target sites, revealing that the RPS5a promoter may affect cytosine base editing efficiencies.

Results

The *lfy* mutant phenotype was induced in the primary generation of transgenic *Arabidopsis*.
To examine and test the efficacy of base editing tools to introduce the non-sense mutation *in planta*, we selected the plant RPS5a promoter and the viral 35S promoter to generate four different construct combinations with two CBEs: BE3 and AIDv2 (Fig. 1A). Target sites for gRNA were chosen to recapitulate previously described loss-of-function *lfy* mutations; targeting site-1 (5′-ATTGC_5AAGAAGAGGAGGAAGAGG-3′) and site-2 (5′-AGAAGC_6AAC_9AGC_{12}AGC_{15}AGAGACGG-3′) was expected to induce nonsense mutations using a CBE at the amino acid positions Q119* (site-1) and Q189*, Q190*, Q191*, or Q192* (site-2), by converting the corresponding cytosine bases to thymine bases (Fig. 1). Once the hygromycin-selected individual T_1 plants reached the reproductive phase, we used the mutant *lfy* phenotype as a proxy for CBE efficiency. For the pCBE-A2 construct, in which the RPS5a promoter controlled CBE AIDv2 expression, carrying an gRNA for target site-1 (Fig. 1A), we observed typical *lfy* mutant phenotypes among the T_1 individuals. Nineteen of 69 selected T_1 plants (28%) had typical *lfy* phenotypes in some branches (Fig. 1C) or in whole plants (Fig. 1D); the *lfy* phenotypes were characterized by abnormal flowers with no distinct petals and no male floral organs. To quantify genotypic changes to the target cytosine, we employed a targeted deep sequencing method using NGS (also refer to the amplicon deep sequencing). We determined the C-to-T conversion rate by calculating the percentage of amplicon reads that had substituted cytosine for thymine at the target site, and used this as an indicator of base editing efficiency. The C_5 (the targeting cytosine position 5 when counting the PAM site as positions 21–23) was queried at the site-1 target by amplicon deep sequencing. DNA that was extracted from branches that displayed the regular wild-type flowers of T_1 plant #9 yielded a C-to-T conversion rate of 24.9%; however, the C-to-T conversion rate was 99.7% for DNA extracted from branches with abnormal flowers or *lfy* mutant phenotypes in the same individual (#9; Fig. 1C). The correlation between the C-to-T conversion rate and the phenotype indicated that partially reduced expression of unmutated *LFY* is sufficient to promote normal floral development, and the loss-of-function phenotype is only observed when mutant copies of the *LFY* gene product are more abundant. In addition, targeted deep sequencing of T_1 plant #14, which was transformed with pCBE-A2 and presented the *lfy* phenotype across all branches, had a 99.8% C-to-T conversion rate at C_5 at site-1 (Fig. 1D). Our data confirmed that the intended CBE-induced mutation was introduced, in which an early stop codon was generated at target site-1 by converting glutamine 119 (CAA:Q) into a stop codon (TAA:Stop) and most importantly pCBE-A2 can induce CBE events in the first generation of transformants, conferring a complete loss-of-function phenotype in some T_1 individuals.

The inheritance of mutant alleles into the secondary generation of transgenic *Arabidopsis*.
To assess the germline transmission of base-edited mutant alleles, we harvested the seeds of T_1 plants that had been transformed with the pCBE-A1 vector (promoter: 35S, CBE: AIDv2) and the pCBE-B2 vector (promoter: RPS5a, CBE: BE3), which targeted site-2 (5′-AGAAGC_6AAC_9AGC_{12}AGC_{15}AGAGACGG-3′). Target site-2 has four cytosine residues which, when converted to thymine, introduce early stop codons and thus nonsense mutations in LFY. We measured the C-to-T conversion rates at each possible nonsense mutation position within site-2 using targeted deep sequencing by NGS from individual T_1 plants that were transformed with pCBE-A1 and pCBE-B2 vectors. The C-to-T conversion rates in six individual T_1 plants by pCBE-A1 ranged from 18.7 to 63.5% at protospacer position 6 (C_6), and from 13.2% to 39.6% at position C_{12}; no significant changes were observed at the two other possible target cytosines, C_9 and C_{15}, which exhibited C-to-T conversion rates of < 1.9% (Fig. 1B, Table S2). The C-to-T conversion rates in T_1 plants transformed by pCBE-B2 showed up to 75.5% at position C_6 but no significant substitutions were found at three other target cytosines (Table S2). We then harvested and sowed seeds from T_1 individuals showed high conversion rates to screen *lfy* mutant
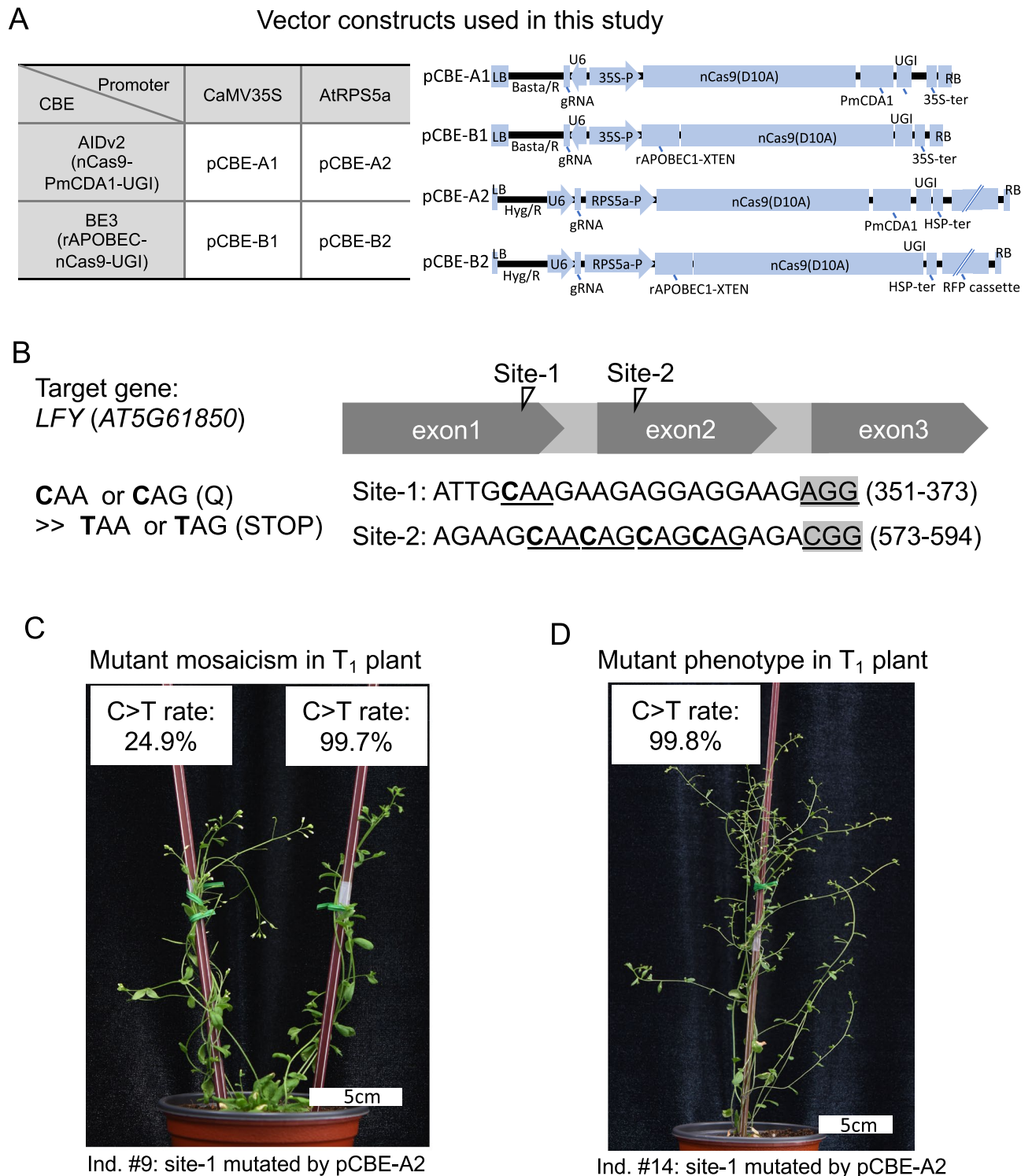
A Vector constructs used in this study

CBE \ Promoter	CaMV35S	AtRPS5a
AIDv2 (nCas9-PmCDA1-UGI)	pCBE-A1	pCBE-A2
BE3 (rAPOBEC-nCas9-UGI)	pCBE-B1	pCBE-B2

B

Target gene: *LFY* (*AT5G61850*)

Site-1   Site-2

exon1   exon2   exon3

CAA or **C**AG (Q) >> **T**AA or **T**AG (STOP)

Site-1: ATTG**C**AAGAAGAGGAGGAAG<u>AGG</u> (351-373)

Site-2: AGAAG**C**AA**C**AG**C**AG**C**AGAGA<u>CGG</u> (573-594)

C Mutant mosaicism in T₁ plant

C>T rate: 24.9%   C>T rate: 99.7%

5cm

Ind. #9: site-1 mutated by pCBE-A2

D Mutant phenotype in T₁ plant

C>T rate: 99.8%

5cm

Ind. #14: site-1 mutated by pCBE-A2

Figure 1. CRISPR/Cas9-mediated cytosine base editing in *Arabidopsis*. (**A**) Summary of constructs used in this study and detailed construct composition for each. *LB* left border, *Basta/R* Basta resistance, *U6* U6 promoter, *UGI* Uracil glycosylase inhibitor, *RB* right border, *HSP-ter* heat shock protein terminator, *Hyg/R* Hygromycin resistance, *RFP* red fluorescent protein. (**B**) All specific cytosine (C) positions targeted in each site are presented in bold and codons are underlined. The nucleotide positions of single guide RNA (gRNA) sites are shown in parentheses. The protospacer adjacent motif (PAM) is underlined. (**C,D**). Examples of T₁ plants showing *lfy* mutant phenotypes. (**C**) An example T₁ plant individual transformed with pCBE-A2 targeting site-1, showing the wild-type phenotype (left) and the *lfy* mutant phenotype (right). C-to-T conversion rates were analyzed using tissues collected from the left or right part for targeted deep sequencing. (**D**) A T₁ individual plant transformed with pCBE-A2 targeting site-1 showing the *lfy* mutant phenotype only. Ind.: individual.
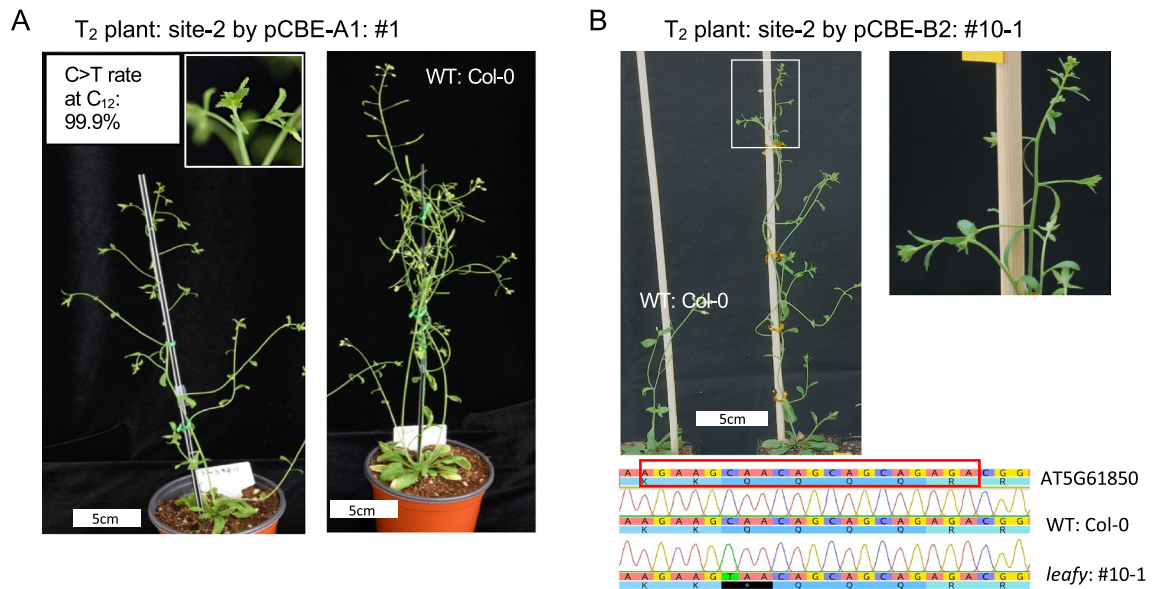
Figure 2. Distinct *lfy* mutant phenotype in T_2 plants induced by CRISPR/Cas9-mediated cytosine base editing in *Arabidopsis*. (**A**) A T_2 plant transformed with pCBE-A1 targeting site-2 displaying the *lfy* mutant phenotype, with the wild-type Col-0 plant for comparison. C-to-T conversion was confirmed for position C_{12}, which had a conversion rate of 99.9% as calculated following targeted deep sequencing analysis. (**B**) A T_2 plant transformed with pCBE-B2 targeting site-2 with the Col-0 plant. The meristem region of inlet was magnified next to the whole plant picture. The complete conversion of C-to-T on the position C_6 was presented by Sanger sequencing result.

phenotypes in T_2 plants for both lines transformed by pCBE-A1 and pCBE-B2. In the next generation, 3 of 102 T_2 individuals by pCBE-A1 presented complete *LFY* knock-out phenotypes, with 99.9% C-to-T conversion rates at position C_{12} only (Fig. 2A); these are from T1 plant with 18.7% C-to-T conversion rate at position C_6 and 13.3% at C_{12}. 2 of 29 T2 individuals from the T1 line with 64.1% C-to-T conversion rate at position C_6 only by pCBE-B2 also showed complete lfy mutant phenotypes (Fig. 2B). To access germline transmission of mutant alleles induced in the T_1 generation we checked whether vector plasmids are remained and active for possible 'de novo' base editing by carrying out amplification of designated four regions with specific PCR primers on T-DNA vector (Fig. S1). No specific PCR products were detected for three T2 individuals by pCBE-A1 nor for one T2 individual by pCBE-B2. The other individual by pCBE-B2 showed PCR bands for three sets of primers but no band for one set targeting on Cas9 region, which indicates base editing activity is inactive (Fig. S1).

To see possible off-target effect on *lfy* mutant phenotypes in these T_2 transgenic lines we listed the potential off-target sites allowing up to three mismatches in target sequence using Cas-Offinder[22] (http://www.rgenome. net). 18 sites were identified; 2 sites containing 2-mismatches, 16 sites containing 3 mismatches with N[A/G]G PAM (5 sites with NGG PAM), and all sites indicated no specific base substitution (Fig. S2).

Construct comparisons with cytosine to thymine conversion rates in T_1 plants. Following validation of *lfy* as a phenotypic marker for functional assays and confirmation of C-to-T conversion activity in our two CBE constructs, we generated more T_1 lines to perform comparative analyses of promoters and CBE systems. All four CBE vector constructs (Fig. 1A) were introduced into *Arabidopsis* via *Agrobacterium*-mediated transformation by floral dipping, and the C-to-T conversion rates of cytosine targets in individual T_1 plants were compared using NGS of the two target site amplicons. We found considerable variation in C-to-T conversion rates between each vector construct. For target site-1, which has only one cytosine target at position 5 (C_5), construct pCBE-A2 (RPS5a-AIDv2) exhibited the highest conversion efficiency of the four constructs tested in the first generation; 32% (22 of 69) of T_1 individuals carrying pCBE-A2 showed > 50% conversion rates at the target cytosine, while no T_1 individuals presented > 50% C-to-T conversion rates following transformation with the other CBE constructs (Fig. 3A). For target site-2, we calculated the C-to-T conversion rates of individuals in which all four potential C-to-T base editing targets were converted. One of 60 (1.3%) T_1 individuals carrying the vector pCBE-A1 (35S-AIDv2) and 10 of 60 T_1 individuals (16.7%) carrying pCBE-B2 (RPS5a-BE3) showed C-to-T conversion rates > 50% (Fig. 3A).

We then categorized the C-to-T conversion rates by two vector construct factors: the 'promoter set' and the 'CBE set'. For the promoter set, we divided all C-to-T conversion rates into two groups. First, we compared ratios obtained from T_1 individuals transformed with vectors containing the RPS5a promoter (pCBE-A2 and pCBE-B2) with those from T_1 individuals carrying the 35S promoter (pCBE-A1 and pCBE-B1). We then grouped C-to-T conversion rates according to the CBE composition; the two groups contained T_1 individuals transformed with vectors encoding AIDv2 (pCBE-A1 and pCBE-A2) or BE3 (pCBE-B1 and pCBE-B2). T_1 individuals carrying RPS5a promoter constructs showed higher C-to-T conversion rates at both sites-1 and -2. A total of 13% of T_1 individuals with the RPS5a promoter (32 of 247) had C-to-T conversion rates > 50%, while only one T_1 individual

A A summary of C-to-T conversion rate in T_1 plants

	Site-1 (ATTGC_5AAGAAGAGGAGGAAG)				Site-2 (AGAAGC_6AAC_9AGC_{12}AGC_{15}AGAGA)			
	No. of T_1 with conversion rate (%)				No. of T_1 with conversion rate (%)			
Constructs	< 10	10 ~ 50	> 50 %	total inds.	< 10	10 ~ 50	> 50 %	total inds.
pCBE-A1	31	5	0	36	71	5	1	77
pCBE-B1	40	0	0	40	35	5	0	40
pCBE-A2	28	19	22	69	76	2	0	78
pCBE-B2	37	3	0	40	26	24	10	60
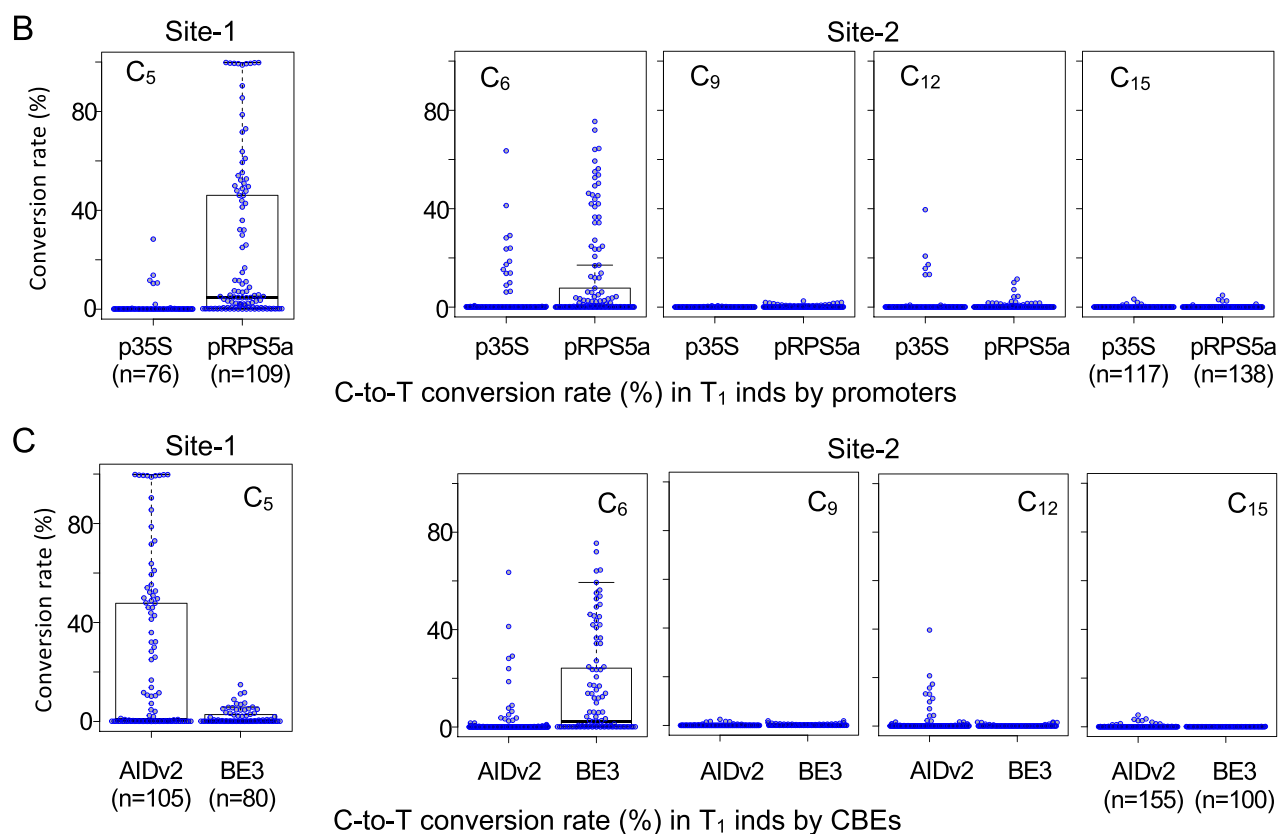
Figure 3. Comparison of the effects of different promoters and base editors on cytosine-to-thymine conversion rates in the first generation of transformed *Arabidopsis* individuals (T_1). (**A**) C-to-T conversion rates calculated from targeted deep sequencing analysis of each T_1 individual are presented. The numbers of individuals with low (< 10%), middle (10–50%), or high (> 50%) conversion rates are indicated. (**B,C**) Comparison of base editing efficiencies with different combinations of promoter and cytosine base editor (CBE) in vector constructs. All C-to-T conversion rates from T_1 individual plants were grouped by vector and the construct used and plotted. (**B**) Comparison of conversion rates grouped by promoter set: CaMV35S versus RPS5a (pCBE-A1 and pCBE-B1) versus (pCBE-A2 and pCBE-B2). (**C**) Comparison of conversion rates grouped by CBE set: AIDv2 versus BE3 (pCBE-A1 and pCBE-A2) versus (pCBE-B1 and pCBE-B2). Each open dot represents the C-to-T conversion rate for each T_1 individual. C-to-T conversion rates were calculated for each targeting cytosine position. *Inds* individuals.

with the 35S promoter (n = 193) had a > 50% conversion rate (Fig. 3B). Sites-1 and -2 displayed different results when the C-to-T conversion rates were grouped by the 'CBE set'. For target site-1, a higher number of T_1 individuals with high conversion rates (> 50%) were observed in AIDv2 CBE transformants (pCBE-A1 and pCBE-A2) compared to BE3 (pCBE-B1 and pCBE-B2) transformants. Conversely, more T_1 individuals transformed with BE3 vectors had high conversion rates at position C_6 of site-2 relative to those transformed with AIDv2 vectors (Fig. 3C). For position C_{12} of site-2, the opposite pattern was observed; the 35S promoter set and the AIDv2 CBE set produced a higher number of T_1 individuals with high C-to-T conversion rates (Fig. 3B,C).

Discussion

Genome editing tools can effectively generate stable transgenic lines in plants; however, predicting how efficiently novel mutations can be introduced by particular CRISPR-Cas9 constructs, and how successfully those mutated loci can be transmitted into the next generation, remain as important challenges[23–25]. Recent studies have identified ways to increase the editing efficiency of CRISPR-Cas9 systems in the T_1 generation, such as using high strength promoters[26–28] and heat treatment[29]. The 35S promoter, which is often used for overexpression experiments in plant research, can be applied to typical CRISPR/Cas9-mediated experiments in which indel mutations are desired, and germline transmission of the mutant alleles is frequently successful[30]. However, 35S may not be the optimal promoter for base editing systems; the efficiency of base editing is generally lower than that of conventional indel-generating Cas9 systems, and lower rates (ca. 25-fold less) of germline transmission have been reported with 35S promoter use compared to other promoters[15,16,18].

Our results provide a clear comparative overview of promoter and CBE system use for base editing, which demonstrates that the RPS5a promoter can possibly increase base editing efficiency in cytosine base editing in *Arabidopsis*. However, we cannot rule out the possibility of stochastic bias in this study due to the small number of gRNA target sites used. Nevertheless, our quantitative mutant read assessment via NGS analyses, which included a large number of T_1 individuals, suggested that RPS5a can be an effective promoter for CBEs. Use of RPS5a showed an increased effect on the CBE efficiency of T_1 individuals in *Arabidopsis* at at least two cytosine sites and had a higher propensity for generating > 50% mutant reads than the 35S promoter. Given that recessive mutations typically would not manifest when cells carry > 50% normal protein levels, our 50% cutoff for genome editing efficiency was biologically relevant; for example, our data showed that tissues carrying < 50% mutant reads mostly displayed normal-looking phenotypes. Thus, promoters that facilitate high CBE mutation rates in the first generation of plants is desirable, as this would enable the function of genes, or particular residues within genes, to be assessed. Although our results might not provide enough number of tests, we can add the idea that the mutation target site sequence would make various base editing outcomes depending on a CBE system; CBE AIDv2 performed better at site-1, whereas CBE BE3 was more efficient at modifying site-2. This suggests that the two CBEs, which harbor different linkers and deaminase properties, may exhibit different preferences for particular target sequences and/or different fidelities depending on the target sequence composition[9,31,32] and also the target base position makes differences[4,6].

Our experimental results prompt us to recommend the use of the RPS5a promoter in vector constructs designed to introduce C-to-T mutations in *Arabidopsis*. We observed that almost 100% C-to-T conversion rates at target site-1 was achievable even in T_1 generation using pCBE-A2 (pRPS5a-AIDv2; Fig. 1C,D). Additionally, a significant number of T_1 plants in which site-2 was targeted showed > 50% C-to-T conversion efficiency when transformed with pCBE-B2 (pRPS5a-BE3; Fig. 3); no plants with > 50% conversion rates were observed following transformation with either construct containing the 35S promoter. We would note that CBE choice may not universally guarantee success, as our quantitative assessment showed a bias when multiple cytosines were targeted simultaneously. Furthermore, CBE choice needs to be considered alongside other parameters, such as target site composition, because the active target window for base editing appears to differ between AIDv2 and BE3. Further investigation is required to determine the target preferences of the two CBEs examined in this study, in which a larger set of target sites should be assessed. However, our results do suggest that AIDv2 may function well for a broader, less specific active window of target base changes, whereas BE3 may be more appropriate for the induction of base substitution mutations specifically at positions 4–6 of the active window. It is also evident that both BE3 and AIDv2 are unsuitable for base changes with an active window size of 1 bp, in other words, without concerning bystander editing; for fine precision edits such as this, Prime Editor could be a more appropriate option[33] or the proper combination of other variants that reduce the active window size[34] and the extended version of PAM limitation such as x-Cas9[35], Cas9-NG[36], near-PAMless SpRY[37].

Another important factor to consider when designing base editing systems is that unwanted indel inductions by nickase Cas9 activity may arise with the use of CBEs. We observed some indel mutations in our amplicon deep sequencing analysis, but the indel ratios and frequencies in T_1 individuals were relatively low; 10 of 185 T_1 individuals harbored indel mutations in site-1 and 11 of 255 T_1 individuals in site-2, giving an indel mutation rate of > 5%. Notably, 8 of 10 and 6 of 11 cases with indel mutation ratios of > 5% resulted from transformation of pCBE-A2 (pRPS5a-AIDv2) targeting site-1 and pCBE-A1 (p35S-AIDv2) targeting site-2, respectively (Table S2).

In conclusion, our investigation into how promoters and CBE systems can affect the efficiency of cytosine base editing experiments in *Arabidopsis* suggested that the RPS5a promoter can improve CBE vector constructs for *Agrobacterium*-mediated transformation by increasing base editing efficiency. However, the efficiency of cytosine base editing may be further affected by the choice of CBE system, target sequence variability, and the active range of the base pair target window. Nonetheless, the use of the RPS5a promoter is likely to be plausible for cytosine base editing in *Arabidopsis*.

Methods

Preparation of vector constructs. All vector constructs used in this study can be found in Fig. 1A. We used the previously published pBAtC[30] binary vector for the backbone to build pCBE-A1 and pCBE-B1, which contain the 35S promoter and nCas9 (D10A)-PmCDA1-UGI (Addgene plasmid number: 79620) and APOBEC1-XTEN-dCas9 (A840H)-UGI (Addgene plasmid number: 73021), respectively. Similarly, pCBE-A2 and pCBE-B2 were constructed by engaging nCas9 (D10A)-PmCDA1-UGI and APOBEC1-XTEN-dCas9 (A840H)-UGI into the previously published pJY-RpCas9 binary vector[18], respectively. To clone guide RNAs in each vector we digested vectors with the restriction enzyme, AarI (ThermoFisher Science, USA) and mixed both annealed product of two oligonucleotides and T4 DNA ligase (New England Biolabs, Ipswich, MA, USA)

for ligation. These ligation products were transferred to *E. coli* DH5alpha and cells containing proper constructs were selected on the media with spectinomycin. All primers used in this study are listed in Table S1.

Target selection and gRNA preparation. The two cytosine base editing target sites were manually selected by scanning the NGG (or CCN for antisense strand searching) along the coding region of LFY gene (At5G61850). The CRISPR vector pCBEs were linearized by the restriction enzyme AarI for 3 h to produce 4 bp overhangs, then purified using the ExpinTM PCR SV mini kit (GeneAll®, Cat# 103–202). Forward and reverse oligos (site-1 and site-2) for each guide sequence were mixed with T4 DNA ligase 10X buffer (New England Bio-Labs®, Cat# B0202S) to 10 μM each and annealed using the following thermocycling program: pre-denaturation for 1 min at 95 °C; 95–25 °C for − 1 °C/min; and hold at 10 °C. The linearized vectors and annealed oligos for guide sequences were ligated and transformed into *E. coli* competent cells as previously described[30]. *E. coli* transformants were selected on spectinomycin; plasmid DNA was isolated using the ExprepTM plasmid SV mini kit (GeneAll®, Cat# 101–102) and validated by Sanger sequencing.

Agrobacterium-mediated transformation. The *Arabidopsis thaliana* Col-0 ecotype was used for *Agrobacterium tumefaciens* GV3101-mediated plant transformation via floral dip[28]. Col-0 plants were grown under long day conditions (16 h light/8 h dark) at 22 °C in a growth room (Koencon, Hanam, South Korea). Light was generated from a 32 W Osram lamp (170 mmol/2 m/s).

Targeted deep sequencing and mutation pattern analysis. Genomic DNA for targeted deep sequencing analysis was extracted from two or more randomly selected leaves from each individual of T_1 or T_2 plants using commercially available plant DNA extraction kit. The on-target sequence was amplified from genomic DNA using target-specific primers (Table S1) that were designed using reference genome sequences and attached with adapter sequences. Multiplexing indices and sequencing adaptors were added by an additional PCR using the protocol supplied from the sequencing company, Macrogen (Seoul, South Korea). High-throughput sequencing was performed using Illumina Mini-seq (San Diego, CA, USA) with the paired-end multiplexed library. Raw reads from paired-end Mini-seq sequencing were joined by the program, 'fastq-join' and implemented in the package, 'eu-util' (http://code.google.com/p/ea-utils). Base substitution rates were calculated using the 'BE-analyzer' on the web-based program (http://rgenome.net) or an in-house python script. The average amplicon read number generated in this study was 11,137 and can be found in Table S2. We also extracted the DNA from several randomly selected leaves of T_2 individuals and amplified the on-target region using the same primer sets used in the targeted deep sequencing. The PCR product was directly sequenced by Macrogen.

Detection of off-target mutations. Three transgenic T2 plants by pCBE-A1 and two by pCBE-B2 on target site-2 were selected for each off-target detection. Potential off-target sites allowing 3 bp mismatches on target sequence were searched using Cas-Offinder[22] based on the Arabidopsis reference genome and amplified with the primer sets listed in supplementary Fig. S2. Purified PCR products were sent for Sanger sequencing (Macrogen, Seoul, South Korea) to identify base substitution.

References

1. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821. https://doi.org/10.1126/science.1225829 (2012).
2. Doudna, J. A. & Charpentier, E. Genome editing The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096. https://doi.org/10.1126/science.1258096 (2014).
3. Sternberg, S. H. & Doudna, J. A. Expanding the biologist's toolkit with CRISPR-Cas9. *Mol. Cell* **58**, 568–574. https://doi.org/10.1016/j.molcel.2015.02.032 (2015).
4. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844. https://doi.org/10.1038/s41587-020-0561-9 (2020).
5. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424. https://doi.org/10.1038/nature17946 (2016).
6. Rees, H. A. & Liu, D. R. Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788. https://doi.org/10.1038/s41576-018-0059-1 (2018).
7. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729. https://doi.org/10.1126/science.aaf8729 (2016).
8. Zong, Y. *et al.* Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440. https://doi.org/10.1038/nbt.3811 (2017).
9. Bharat, S. S., Li, S., Li, J., Yan, L. & Xia, L. Base editing in plants: Current status and challenges. *Crop J.* **8**, 384–395. https://doi.org/10.1016/j.cj.2019.10.002 (2020).
10. Lu, Y. & Zhu, J. K. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 523–525. https://doi.org/10.1016/j.molp.2016.11.013 (2017).
11. Shimatani, Z. *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443. https://doi.org/10.1038/nbt.3833 (2017).
12. Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 526–529. https://doi.org/10.1016/j.molp.2016.12.001 (2017).

13. Wu, J. *et al.* Engineering herbicide-resistant oilseed rape by CRISPR/Cas9-mediated cytosine base-editing. *Plant Biotechnol. J.* **18**, 1857–1859. https://doi.org/10.1111/pbi.13368 (2020).
14. Cheng, H. *et al.* Base editing with high efficiency in allotetraploid oilseed rape by A3A-PBE system. *Plant Biotechnol. J.* **19**, 87–97. https://doi.org/10.1111/pbi.13444 (2021).
15. Castel, B., Tomlinson, L., Locci, F., Yang, Y. & Jones, J. D. G. Optimization of T-DNA architecture for Cas9-mediated mutagenesis in Arabidopsis. *PLoS ONE* **14**, e0204778. https://doi.org/10.1371/journal.pone.0204778 (2019).
16. Ordon, J. *et al.* Optimized Cas9 expression systems for highly efficient Arabidopsis genome editing facilitate isolation of complex alleles in a single generation. *Funct. Integr. Genom.* **20**, 151–162. https://doi.org/10.1007/s10142-019-00665-4 (2020).
17. Shockey, J. Gene editing in plants: Assessing the variables through a simplified case study. *Plant Mol. Biol.* **103**, 75–89. https://doi.org/10.1007/s11103-020-00976-2 (2020).
18. Kang, B. C. *et al.* Precision genome engineering through adenine base editing in plants. *Nat. Plants* **4**, 427–431. https://doi.org/10.1038/s41477-018-0178-x (2018).
19. Parcy, F., Bomblies, K. & Weigel, D. Interaction of LEAFY, AGAMOUS and TERMINAL FLOWER1 in maintaining floral meristem identity in Arabidopsis. *Development* **129**, 2519–2527 (2002).
20. Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F. & Meyerowitz, E. M. LEAFY controls floral meristem identity in Arabidopsis. *Cell* **69**, 843–859. https://doi.org/10.1016/0092-8674(92)90295-n (1992).
21. Huala, E. & Sussex, I. M. LEAFY interacts with floral homeotic genes to regulate Arabidopsis floral development. *Plant Cell* **4**, 901–913. https://doi.org/10.1105/tpc.4.8.901 (1992).
22. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475. https://doi.org/10.1093/bioinformatics/btu048 (2014).
23. Feng, Z. *et al.* Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. U S A* **111**, 4632–4637. https://doi.org/10.1073/pnas.1400822111 (2014).
24. Mao, Y. *et al.* Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol. J.* **14**, 519–532. https://doi.org/10.1111/pbi.12468 (2016).
25. Mao, Y., Botella, J. R. & Zhu, J. K. Heritability of targeted gene modifications induced by plant-optimized CRISPR systems. *Cell Mol Life Sci.* **74**, 1075–1093. https://doi.org/10.1007/s00018-016-2380-1 (2017).
26. Wang, Z. P. *et al.* Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol.* **16**, 144. https://doi.org/10.1186/s13059-015-0715-0 (2015).
27. Tsutsui, H. & Higashiyama, T. pKAMA-ITACHI vectors for highly efficient CRISPR/Cas9-mediated gene knockout in *Arabidopsis thaliana*. *Plant Cell Physiol.* **58**, 46–56. https://doi.org/10.1093/pcp/pcw191 (2017).
28. Feng, Z. *et al.* A highly efficient cell division-specific CRISPR/Cas9 system generates homozygous mutants for multiple genes in Arabidopsis. *Int. J. Mol. Sci.* **19**, 3925. https://doi.org/10.3390/ijms19123925 (2018).
29. LeBlanc, C. *et al.* Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *Plant J.* **93**, 377–386. https://doi.org/10.1111/tpj.13782 (2018).
30. Kim, H. *et al.* A simple, flexible and high-throughput cloning system for plant genome editing via CRISPR-Cas system. *J. Integr. Plant Biol.* **58**, 705–712. https://doi.org/10.1111/jipb.12474 (2016).
31. Wang, Y. *et al.* Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. *Genome Biol.* **20**, 218. https://doi.org/10.1186/s13059-019-1839-4 (2019).
32. Arbab, M. *et al.* Determinants of base editing outcomes from target library analysis and machine learning. *Cell* **182**, 463–480. https://doi.org/10.1016/j.cell.2020.05.037 (2020).
33. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157. https://doi.org/10.1038/s41586-019-1711-4 (2019).
34. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376. https://doi.org/10.1038/nbt.3803 (2017).
35. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63. https://doi.org/10.1038/nature26155 (2018).
36. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262. https://doi.org/10.1126/science.aas9129 (2018).
37. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290. https://doi.org/10.1126/science.aba8853 (2020).

Acknowledgements

Author contributions

S.-T.K., J.-Y.Y. and J.-S.K. conceived the project; S.-T.K., M.C., and J.-Y.Y. designed and performed the experiments; S.T.K., M.C. and J.H.K. analyzed the data; and S.-T.K. wrote the majority of the manuscript with input from all other authors. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-87669-y.

Correspondence and requests for materials should be addressed to J.-S.K. or S.-T.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.