Engineering Cas9 for more predictable and efficient genome editing

Kwok Ling Yi Samantha
Magdalene College

Declaration

I understand the University's definition of plagiarism. I declare that, in accordance with Discipline regulation 6, this dissertation is entirely my own work except where otherwise stated, either in the form of citation of published work, or acknowledgement of the source of any unpublished material.

Kwok Ling Yi Samantha
28th April 2020

Engineering Cas9 for more predictable and efficient genome editing

Candidate number: 6557W

Official supervisor: Prof. Jussi Taipale

Day-to-day supervisor: Dr. Otto Kauko

Word count: 4998

Summary

Prime editing is a 'search-and-replace' genome editing technology. It works by first nicking the protospacer adjacent motif (PAM)-containing strand using a Cas9 nickase and then adding the desired edit using a reverse transcriptase fused to the nickase. Unlike CRISPR-Cas9 genome editing based upon homology-directed repair, prime editing does not generate double-stranded breaks or utilise donor DNA templates. In the original design, prime editor 2 (PE2) (Anzalone A, Randolph P, Davis J, Souza A, Koblan L, Levy J, Chen P, Wilson C, Newby G, Raguram A & Liu D (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576:** 149–157) was directed by a prime editing guide RNA (pegRNA) to cleave the PAM strand and add on the edit. Then, using a single guide RNA to direct the Cas9 nickase, the other strand that does not carry the edit is cleaved to bias DNA repair to permanently install the edit. Generating a prime editor containing a nuclease with time-separated cleavage would allow directing the whole genome editing process with one pegRNA. This could lead to more efficient and predictable genome editing and potentially enable pooled functional genomic screening. In this project, the Cas9 nickase in PE2 was reverted back to Cas9 nuclease and further mutations were made. The mutated PE2 was expressed in HEK293T cells together with pegRNA. Reparation of broken green fluorescent protein was used to determine editing efficiency and laser scanning cytometry was used to detect, image, and quantify cells. Here, we report a rationally engineered Cas9 variant (D839N) that improves editing efficiency in PE2. Additionally, we found that a prime editor with both endonuclease domains catalytically active can still edit and can be used in genome editing to introduce large deletions. Furthermore, we found 22 other mutations which result in editing efficiency higher than the abovementioned double strand-cutting prime editor but lower than PE2. These mutations can be combined to potentially give rise to a prime editor with improved editing efficiency.

Note: pegRNA assembly was done by Dr. Kauko due to time constraints.

Table of Contents

Abbreviations

BFP, blue fluorescent protein;
CRISPR-Cas9, clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9;
DSB, double-stranded break;
GFP, green fluorescent protein;
HDR, homology-directed repair;
PAM, protospacer adjacent motif;
PBS, primer-binding site;
PE2, prime editor 2;
pegRNA, prime editing guide RNA;
Pre-crRNA, precursor CRISPR RNA;
sgRNA, single guide RNA;
tracrRNA, trans-activating crRNA;

Introduction

Clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9 (CRISPR-Cas9) is an RNA-guided DNA targeting platform derived from bacterial and archaeal defence systems that can be used for genome editing (Jinek et al, 2012). The programmable nuclease generates double-stranded DNA breaks (DSBs) that can give rise to insertions and deletions at target sites (Cong et al, 2013; Mali et al, 2013; Jinek et al, 2012).

CRISPR-Cas systems are made of Cas genes and a CRISPR array. The CRISPR array has spacers, which are foreign genetic material-targeting sequences, interspersed with identical repeats (Wiedenheft et al, 2012). Protospacers, which are foreign sequence fragments, are added onto the proximal end of the CRISPR array (Bhaya et al, 2011; Terns & Terns, 2011). This repeat-spacer element is then transcribed into precursor CRISPR RNA (pre-crRNA). In type II systems, a trans-activating crRNA (tracrRNA) complementary to repeat sequences of pre-crRNA activates processing of the pre-crRNA by RNase III, yielding crRNAs (Deltcheva et al, 2011; Gottesman, 2011).

A two-RNA structure, formed by the base pairing of crRNA to tracrRNA, guides Cas9 to introduce DSBs to the invading DNA (Brouns et al, 2008). Cleavage occurs where the crRNA is complementary to the target protospacer DNA. Additionally, the recognition site has to be upstream of a Protospacer Adjacent Motif (PAM). For Streptococcus pyogenes Cas9, the PAM sequence is 5'-NGG-3'. The Cas9 HNH domain cleaves the complementary strand and the RuvC-like domain cleaves the non-complementary strand (Jinek et al., 2012). The tracrRNA-crRNA complex can be engineered as a single guide RNA (sgRNA) to direct cleavage in genome editing (Fig. 1).

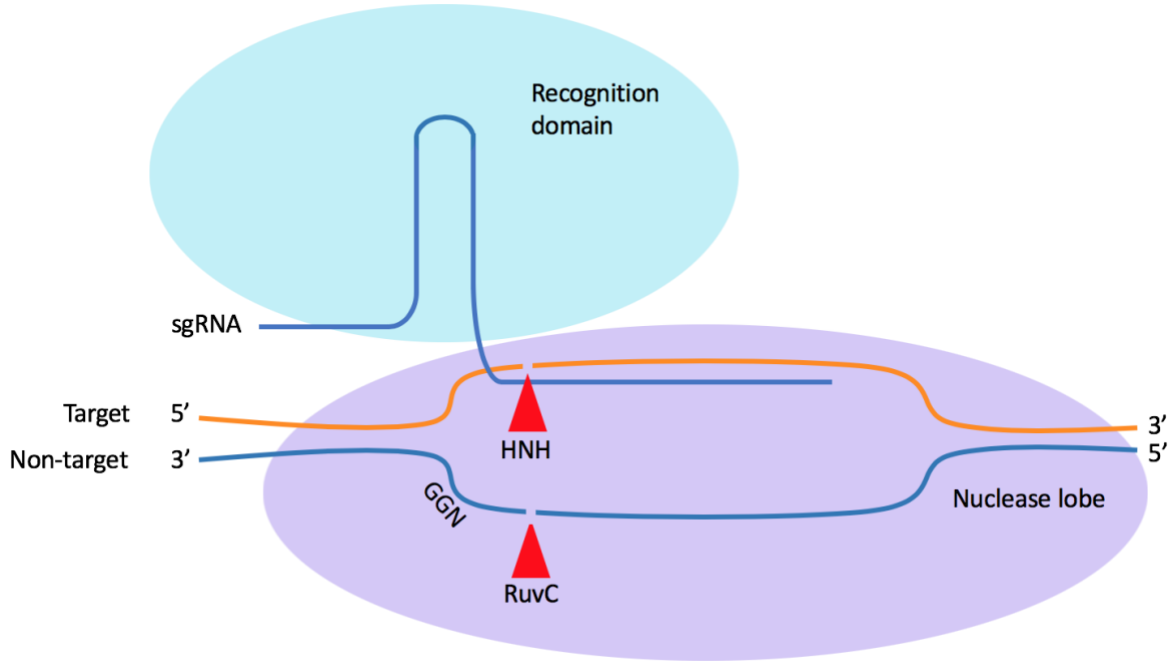

Figure 1. Schematic of CRISPR-Cas9 nuclease structure together with sgRNA at target locus. Schematic illustration of the CRISPR-Cas9 nuclease being recruited to the target DNA site by an sgRNA. The recognition domain interacts with the sgRNA and the nuclease lobe interacts with the target DNA and PAM (sequence 5'-NGG-3'). Two endonuclease domains, HNH and RuvC, cleave the target and non-target strands respectively, creating a DSB.

There are drawbacks with CRISPR-Cas9 genome editing. DSBs can cause negative effects such as mixtures of different products, translocations, and p53 activation (Haapaniemi et al, 2018; Ihry et al, 2018). This is not optimal for reverting pathogenic alleles which are generated by specific insertions, deletions, or substitutions, as precise editing is required (Anzalone et al, 2019).

Another drawback is related to non-homologous end joining. Homology-directed repair (HDR) stimulated by DSBs is used to install precise DNA changes by using donor DNA repair templates (Rouet et al, 1994). However, the DNA damage response leads to activation of competing repair processes and non-homologous end joining of DSBs is the most frequent outcome. This usually gives rise to an excess of insertions and deletions by-products. Additionally, it is not efficient enough in most therapeutically relevant cell types (Cox et al, 2015; Chapman et al, 2012).

Prime editing is a 'search-and-replace' genome editing technology that does not require generating DSBs or utilising donor DNA templates (Anzalone, 2019). It has significantly lower off-

target activity than Cas9 and lower levels of byproducts alongside greater or similar efficiency when compared to Cas9-mediated HDR.

The prime editor is made of a reverse transcriptase fused to an RNA-programmable nickase. As mentioned previously, the targeting specificity of Cas9 can come from an sgRNA. In prime editing, a prime editing guide RNA (pegRNA) both specifies the target DNA and encodes new genetic information to be added. It comprises a protospacer, constant backbone, and a 3' extension which acts as a template for reverse transcription primed by the cleaved DNA strand. The PAM-containing strand is first nicked at the target site, exposing a 3'-hydroxyl which then primes the reverse transcription of the template on the pegRNA onto the target site (Fig. 2).

This results in an intermediate containing two single-stranded DNA flaps. The 5' flap has the unedited sequence whereas the 3' flap contains the edited sequence. 5' flaps are preferentially excised by structure-specific endonucleases like FEN122 as well as 5' exonucleases like EXO123. Along with 5' flap excision, 3' flap ligation aids in incorporating the edited DNA strand. This generates double-stranded DNA with one edited and one unedited strand. DNA repair can be biased to copy the edits from the edited strand to the unedited strand by nicking the unedited strand. Anzalone et al, 2019 achieves this using the nickase in the prime editor guided by an sgRNA targeting the strand complementary to the one targeted by pegRNA.

In order to minimise DSBs, the unedited strand needs to be nicked only after resolving the edited strand flap. Hence, the sgRNA has spacers that only match the edited strand. Due to imperfect complementarity between the spacer and the unedited allele, sgRNA-directed nicking is disfavoured until after the PAM-containing strand has been edited.

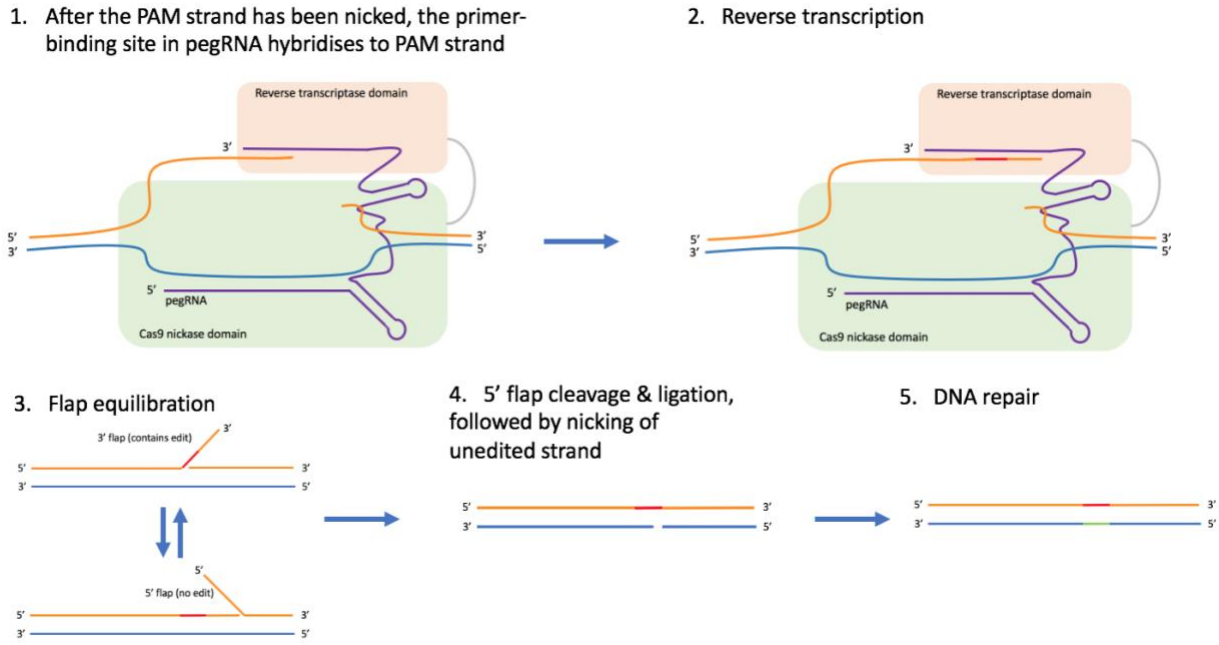

Figure 2. Overview of prime editing.

The PAM-containing strand is nicked by the Cas9 nickase domain. The resulting 3' end primes reverse transcription using the reverse transcription template of the pegRNA. Flap equilibration, 5' flap removal, and ligation results in a heteroduplex of one edited and one unedited strand. Nicking of the unedited strand biases DNA repair to that strand, resulting in stably edited DNA in both strands.

The aim of the project is to alter the kinetics of DNA cleavage in the HNH domain to slow it down, thus achieving time separation between the cleavage of the DNA strands. This way, the RuvC domain first cleaves the PAM-containing strand and allows addition of the desired edit through reverse transcription. Afterwards, the HNH domain cleaves the unedited strand to direct DNA repair to the strand. This allows the usage of a single pegRNA to direct the whole process of genome editing without the usage of another sgRNA to nick the unedited strand. This is in contrast with the original prime editor 2 (PE2) which contains an inactive HNH domain.

By using only one pegRNA instead of both a pegRNA and a sgRNA, the simplification of the prime editing process will allow more efficient and predictable editing. Cells only need to be transfected with the pegRNA and the prime editor constructs and not with an additional single guide RNA construct. The time separation between the cleavage of the DNA strands allows biasing of DNA repair to copy the edits from the edited strand to the unedited strand without creating DSBs. This increases the efficiency of genome editing.

An important benefit of only using one pegRNA is that this allows pooled functional genomic screening, where a library of pegRNAs can be introduced into stable prime editor-expressing cells. These cells can then be placed under selection pressure and have the pegRNA abundance over time quantified by next-generation sequencing, showing the drop-out or enrichment of cells (Chan et al, 2019).

The approach was to design mutations that may slow down the HNH domain. Mutations that either increase or decrease the editing efficiencies were then screened for. This approach was chosen as both types of mutants may be useful. It is possible that mutants with higher editing efficiencies have achieved the aim of slowing down the HNH domain sufficiently to achieve time separation between the cleavage of the DNA strands. Mutants with lower editing efficiencies may also slow down the HNH domain, but insufficiently. However, when these mutations are combined, they could form a prime editor with higher editing efficiency.

Therefore, the structure of Cas9 was studied to design mutations on PE2 (Anzalone et al, 2019). For example, amino acids in the active site were replaced with amino acids of similar character (Biertumpfel, Yang, & Suck, 2007). Additional mutations include those which impair or weaken the interaction of Cas9 with the DNA strand, those which affect linkers involved in HNH repositioning, and those which impact HNH interactions with other domains (Zuo & Liu, 2017).

Cloning and site-directed mutagenesis were attempted using different methods like recombination-based mutagenesis (Watson & García-Nafría, 2019) as well as the Q5 and Quikchange II kits. Ultimately, Quikchange II was selected for large-scale mutagenesis. The prime editor and the pegRNA constructs were then transfected into HEK 293T cells using FuGENE® HD. The HEK 293T cells had genomic insertion of DNA encoding mutated green fluorescent protein (GFP) to be repaired by the prime editor as well as blue fluorescent protein (BFP) for cell quantification. Monitoring editing of the GFP reporter gene was done using laser scanning imaging cytometry.

Here, we report a rationally engineered Cas9 variant (D839N) that improves editing efficiency in PE2 (Anzalone, 2019). Additionally, we found that a prime editor with both endonuclease domains catalytically active can still edit and can be used in genome editing to introduce large deletions. Furthermore, we also found 22 other mutations which result in editing efficiency higher than the abovementioned double strand-cutting prime editor but lower than PE2. These mutations can be combined to potentially give rise to a prime editor with improved editing efficiency.

Results and Discussion

Selecting PEG RNAs

The HEK293T BFP-GFP reporter cell line contains GFP with 5 substitutions that make it non-fluorescent. Screening the editing efficiencies of mutated PE2 with both a more efficient pegRNA and a less efficient pegRNA would allow better comparison as to whether the mutated PE2 alters editing efficiency. This is because with the high efficiency pegRNA, the assay may start to get saturated and it would be difficult to see the improvement if PE2 is very efficient. With the less efficient pegRNA, it may be difficult to find mutants with lower efficiency. Hence, pegRNA1 (1/1, 1/2, 1/3), pegRNA2 (2/1, 2/2, 2/3), and pegRNA3 (3/1, and 3/2) (Fig. 3A) were tested. pegRNAs 2/3 and 3/2 were selected to use in screening as the more efficient and less efficient pegRNAs respectively. Fig. 4 shows some images obtained from Acumen® Cellista used to calculate the percentage of edited cells.

Mutating the PAM sequence seems to increase the editing efficiency. Overall, pegRNA2 which inactivates the PAM sequence through a base substitution mutation shows greater editing efficiencies as compared to pegRNA1 and pegRNA3/1 which do not mutate the PAM sequence (Fig. 3B). pegRNA3/2 which inactivates the PAM sequence is has greater editing efficiencies than pegRNA3/1 (Fig. 3B). This could be because further editing is prevented with the inactivation of the PAM sequence, resulting in more stable repairing of the GFP.

While both pegRNA1 and pegRNA3/1 do not mutate the PAM sequence, pegRNA1 is consistently more efficient than pegRNA3/1 (Fig. 3B). This might potentially be due to the lower G/C content of the priming-binding site of pegRNA3/1 as compared to pegRNA1, given both pegRNAs have a primer-binding site (PBS) of 13 nucleotides. Anzalone et al, 2019 had a similar observation that priming regions with lower G/C content usually need longer PBS sequences. This agrees with energetic requirements of binding of the nicked DNA strand to the PBS of the pegRNA.

Surprisingly, pegRNA2 with the shortest homology (2/3) has the highest editing efficiency (Fig. 3B). This could potentially be attributed to a lesser secondary structure formation and suggests that the reverse transcription template secondary structure also influences editing activity. This is consistent with the suggestion by Anzalone et al, 2019 that given no PBS length or G/C content was directly indicative of editing efficiency, other factors like reverse transcription template secondary structure also influence editing activity. An alternative explanation could be because the edit by pegRNA2 is closer to the cleavage site.

Additionally, it was noted that accurate determination of the edited fraction would require re-plating the cells after four days. Hence, the screening protocol was adapted. In the screening of mutated PE2, four days after transfection, cells were trypsinised, resuspended, and diluted with media to 25% of the original cell density. Laser scanning cytometry followed immediately after re-plating.

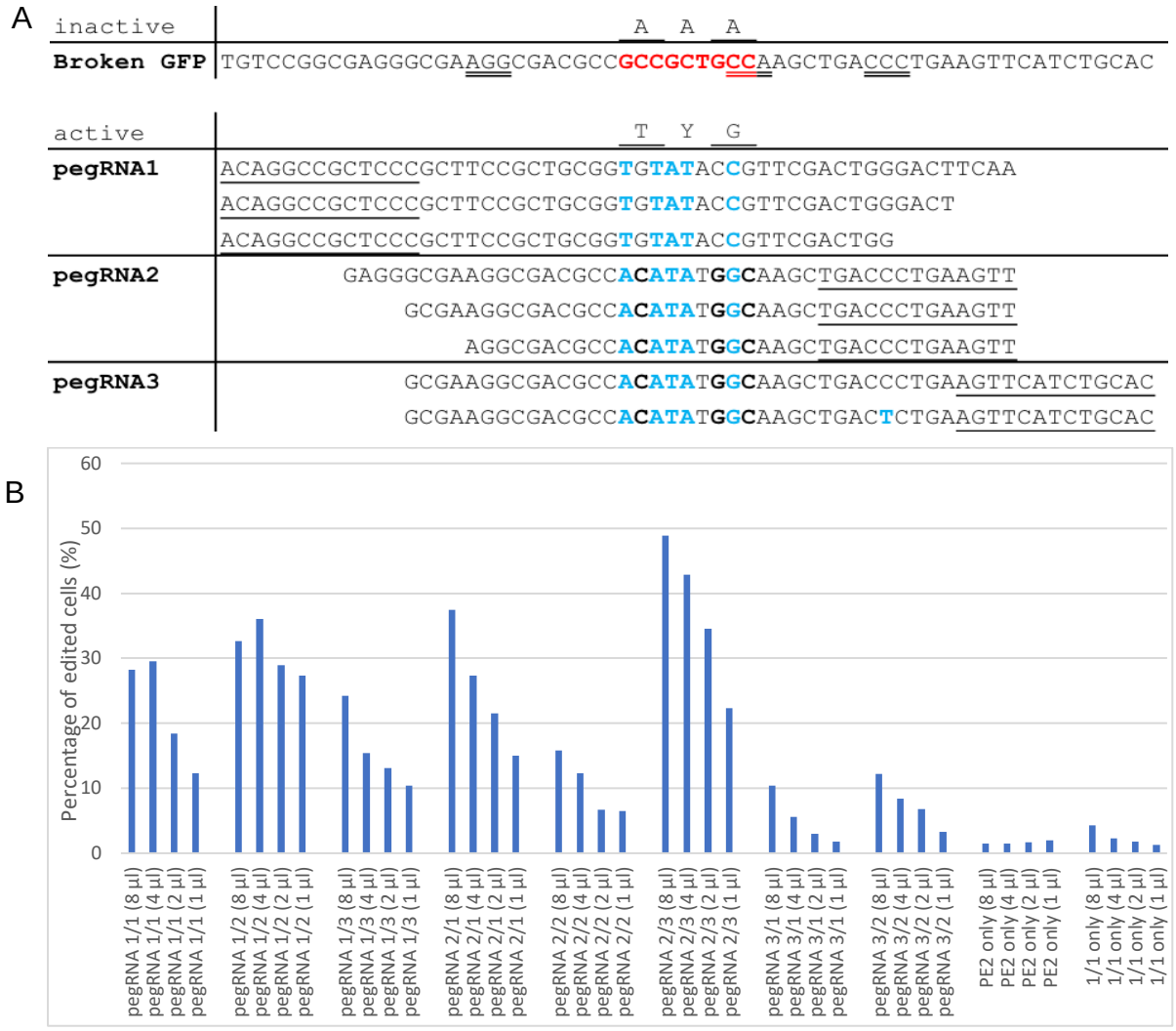

Figure 3. Comparison of the editing efficiency of different PEG RNAs.

A Sequences involved in repairing the broken GFP. Top: Partial sequence of the broken GFP. 3 Inactivating alanine mutations are marked with red and PAM sequences are marked with double underline. Bottom: pegRNAs designed for each PAM (numbered from left to right). From top to bottom, the sequences are pegRNA 1/1, 1/2, 1/3, 2/1, 2/2, 2/3, 3/1, 3/2. The underlined part is complementary to the DNA strand that primes the reverse transcription reaction. The mutations are highlighted with blue. Only the 3' extension of each pegRNA is shown.

B 3 days after transfection, laser scanning cytometry (Acumen® Cellista) was used to quantify the percentage of edited cells (%) for each PEG RNA. PE2 only and 1/1 only were used as negative controls to show that PE2 alone or the pegRNA alone does not lead to significant editing. The volumes in brackets refer to the volume of transfection mixture added to 100 µl of media in each well.

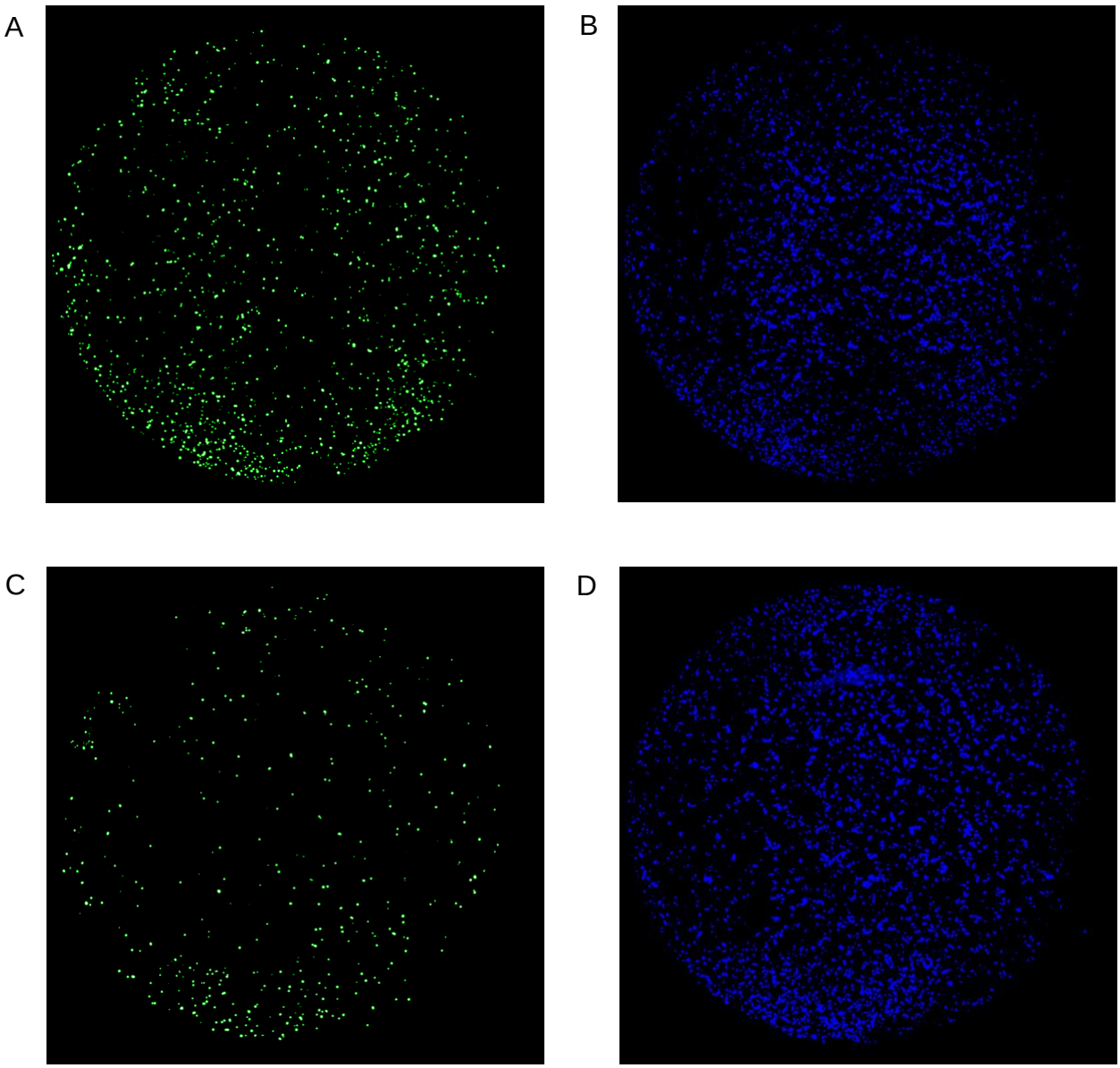

Figure 4. Acumen® Cellista images comparing pegRNA 2/3 and 3/2

A HEK293T cells transfected with pegRNA 2/3 and PE2. Emission was measured at 488 nm for GFP. 42% of cells were edited.

B HEK293T cells transfected with pegRNA 2/3 and PE2. Emission was measured at 405 nm for BFP.

C HEK293T cells transfected with pegRNA 3/2 and PE2. Emission was measured at 488 nm for GFP. 14% of cells were edited.

D HEK293T cells transfected with pegRNA 3/2 and PE2. Emission was measured at 405 nm for BFP.

Transfection volume optimisation

According to Wang et al, 2018, while FuGENE® HD shows a very high transfection efficacy of 41% for HEK293 cells, it also displays strong toxicity of 66% to the cells. Cytotoxicity is an important concern as toxicity may non-specifically affect genes and thus influence experimental results. Therefore, to evaluate the effect of the volume of transfection mixture on cell viability or proliferation, different volumes of transfection mixture (8 µl, 4 µl, 2 µl, 1 µl) prepared according to FuGENE® HD manufacturer's protocol were added to 100 µl of media in each well. Three days after transfection, laser scanning cytometry was used to quantify the number of cells.

While the percentage of edited cells increases with the volume of transfection mixture added (Fig. 3B), the cell viability or proliferation drops at 8 µl of transfection mixture added (Fig. 5). The addition of 8 µl of transfection mixture containing pegRNA2/3 and 840A resulted in only 78% of the number of cells as compared to having no transfection mixture added. This might be because the transfection reagents affected cell viability or proliferation. Hence, 4 µl of transfection mixture was used in subsequent experiments to maximise the transfection efficacy while limiting toxicity.

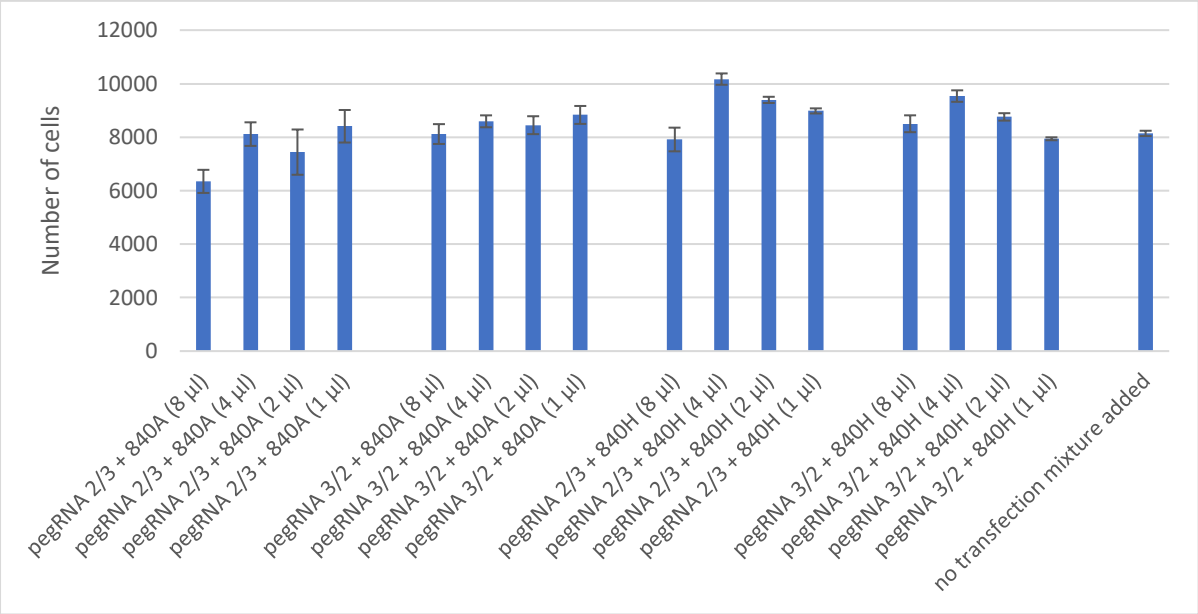

Figure 5. Effect of transfection mixture volume on cell viability.

3 days after transfection, laser scanning cytometry (acumen Cellista) was used to quantify the number of cells. The volumes in brackets refer to the volume of transfection mixture added to 100 µl of media in each well. 840A refers to the original PE2 containing the Cas9 nickase whereas 840H refers to PE2 with the Cas9 reverted back to a nuclease. Standard error bars were obtained by dividing standard deviation by the square root of the number of samples (n = 6 for 840H, n = 2 for 840A, and n = 16 for no transfection mixture added).

Reverting Cas9 nickase (840A) to Cas9 nuclease (840H)

Results from Fig. 6A were obtained 2.5 days after transfection. When the cells were cultured for 5 days after transfection, editing efficiencies were very high even for Cas9 nuclease (840H), at 19% of Cas9 nickase (840A) for PEG RNA 3/2 and 75% of Cas9 nickase (840A) for PEG RNA 2/3. This can be attributed to continuous expression of the plasmid. This shows even with the Cas9 nuclease instead of nickase, PE2 can still edit DNA. An important implication is that PE2 with Cas9 nuclease can be utilised to introduce large deletion mutations (Fig. 6B).

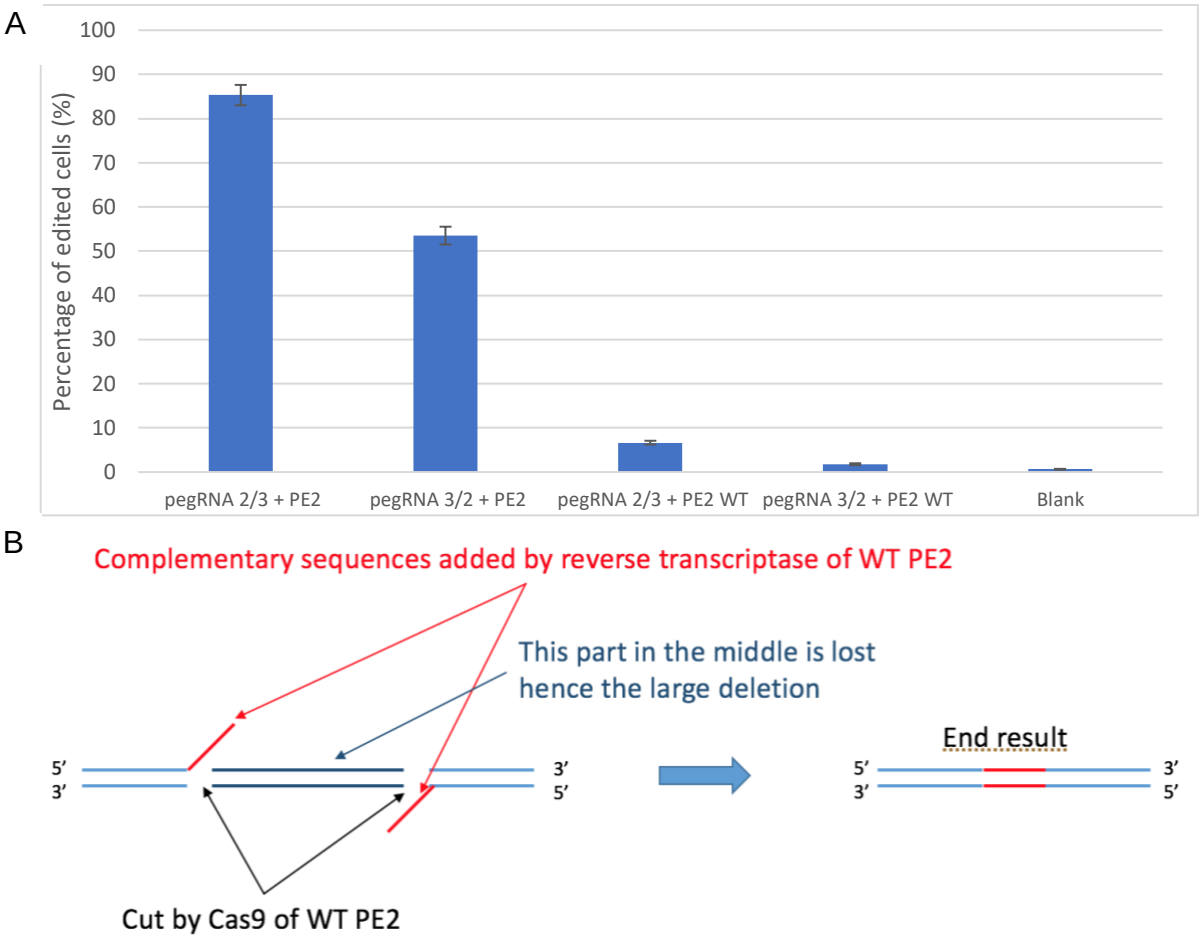

Figure 6. Reverting Cas9 nickase (840A) to Cas9 nuclease (840H) in PE2

A 2.5 days after transfection, laser scanning cytometry (Acumen® Cellista) was used to quantify the percentage of edited cells (%). Cas9 nickase (840A) is labelled as PE2 and Cas9 nuclease (840H) is labelled as PE2 WT. Standard error bars were obtained by dividing standard deviation by the square root of the number of samples (n = 6 for 840H, n = 2 for 840A, and n = 12 for blank).

B Schematic illustration of how PE2 containing Cas9 nuclease can be used to introduce large deletions. The region in dark blue is that to be deleted and the regions in red are the complementary sequences added by reverse transcriptase of WT PE2.

Single PE2 mutants purified with Miniprep

Further mutations were made using PE2 WT which contains the Cas9 nuclease (840H) as a template. Laser scanning cytometry showed that the D839N and D839Q mutants edited a significantly higher percentage of cells as compared to the PE2 with the Cas9 nickase (Fig. 7A). This may indicate that the mutation slowed down the HNH domain sufficiently to achieve time separation between the cleavage of the DNA strands, thus leading to efficient editing. Alternatively, they may still have an inactive HNH but retain Mg^{2+} and thus bind DNA more strongly, or they may stabilise the protein more compared to the PE2 with the inactive HNH domain.

Previous biochemical studies aid in suggesting explanations behind the effects of the D839N mutation. They have proposed that the HNH domain utilises a one-metal-ion hydrolysis mechanism to break the phosphodiester bond linking bases 3 and 4 upstream of PAM (Chen, 2014; Jinek et al, 2012; Gasiunas et al, 2012; Biertümpfel et al., 2007; Li et al., 2003). Asp839 of the Cas9 HNH domain corresponds to Asp40 of Endo VII which coordinates Mg^{2+} (Biertümpfel et al., 2007). The mutation of Asp839 to Asn839 may weaken the interaction with Mg^{2+} and hence affects the ion's coordination with the oxygen atoms of the phosphate group, thus slowing down the phosphodiester bond cleavage. Similarly to the D839N mutation, the D839Q mutation weakens the amino acid's interaction with Mg^{2+} as a negatively-charged carboxyl group is replaced with an uncharged amide group.

Mutations N863D and D839E resulted in a significantly less efficient PE2 (Fig. 7). This suggests these mutations may also slow down the HNH domain, but insufficiently. However, when these mutations are combined, they could lead to a prime editor with higher editing efficiency.

Regarding the D839E mutation, while both aspartic acid and glutamic acid both contain a carboxyl group, the position is different. The strength of interaction of the carboxyl group with Mg^{2+} may be affected, or the positioning of Mg^{2+} may be changed. The ion's coordination with the oxygen atoms of the phosphate group is affected, thus slowing down the phosphodiester bond cleavage.

Previous studies have shown that the N863A mutant functions as a nickase, indicating that Asn863 participates in catalysis. Specifically, Asn863 coordinates Mg^{2+} (Nishimasu et al, 2014). As an uncharged amide group is replaced with a negatively-charged carboxyl group with an N863D mutation, coordination with Mg^{2+} may be affected, which could slow down the phosphodiester bond cleavage.

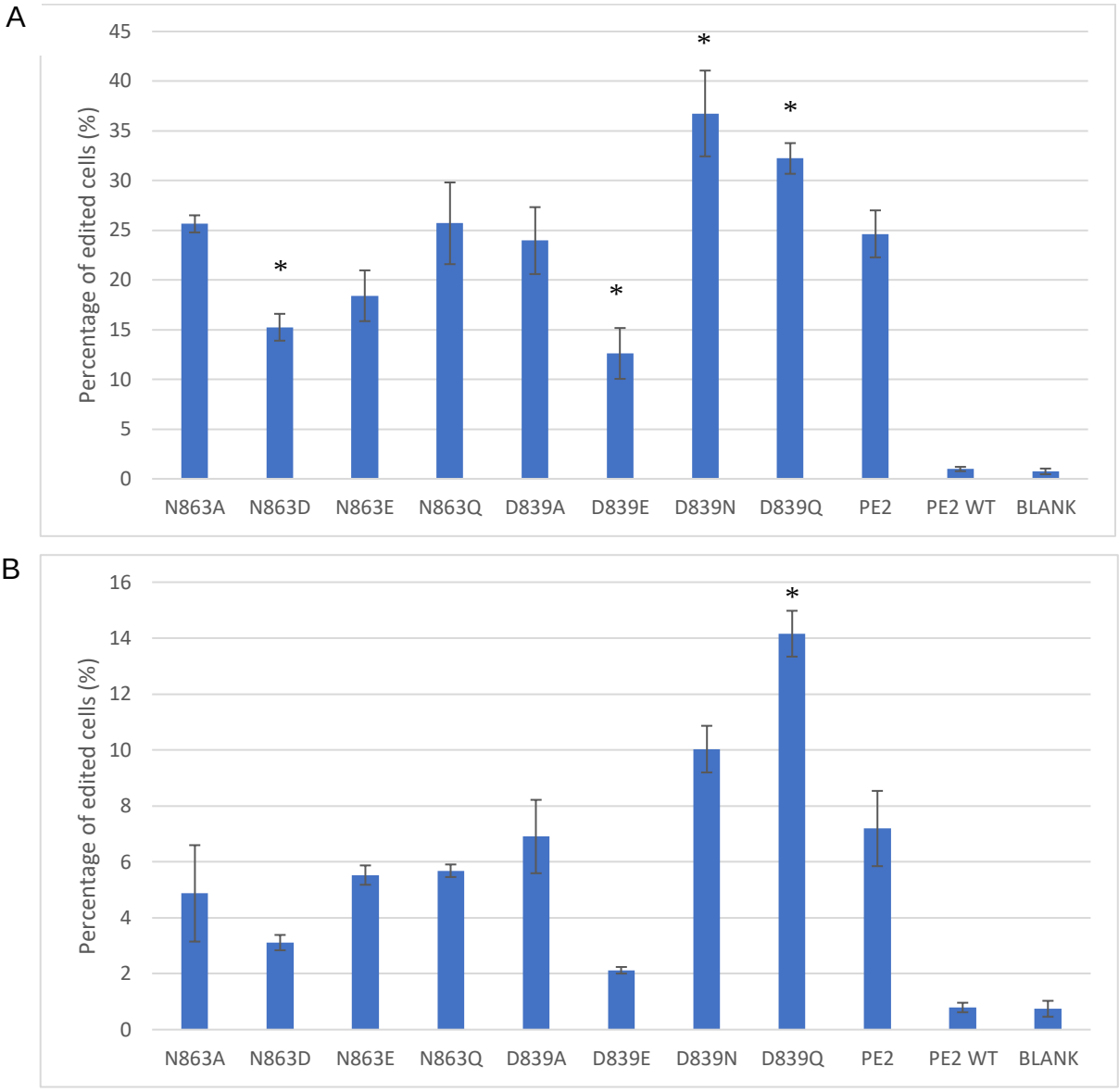

Figure 7. Genome editing efficiencies of different PE2 mutants purified with miniprep

Laser scanning cytometry (acumen Cellista) was used to quantify the percentage of edited cells (%). Cas9 nickase (840A) is labelled as PE2 and Cas9 nuclease (840H) is labelled as PE2 WT. Standard error bars were obtained by dividing standard deviation by the square root of the number of samples (n = 4 for mutants, n = 5 for PE2 and PE2 WT, n = 12 for blank). The stars show significance as compared to PE2. Significance was determined using a two-tailed, two-sample t-test.

A The cells were transfected with the more efficient PEG RNA 2/3 along with the mutated PE2.

B The cells were transfected with the less efficient PEG RNA 3/2 along with the mutated PE2.

Single PE2 mutants purified with Maxiprep

To further confirm the previous results (Fig. 7), plasmids were prepared again from bacterial liquid culture and purified with Plasmid Maxiprep Kit (Qiagen) instead of Plasmid Miniprep Kit (Qiagen) as used previously. This is because Maxiprep results in better quality plasmid that is endotoxin-free while Miniprep does not remove endotoxin. Endotoxin has been shown to decrease cell viability (Cotten et al, 1994) and transfection efficiency (Weber et al, 1995).

Results still showed that mutations N863D and D839E resulted in a significantly less efficient PE2. (Fig. 8) However, surprisingly mutation D839Q also resulted in a significantly less efficient PE2 when pegRNA 2/3 was used (Fig. 8A). While the D839N mutant still resulted in a more efficient PE2, the difference was not statistically significant (p-value of 0.2399 for pegRNA 2/3 and 0.0884 for pegRNA 3/2) (Fig. 8). In future experiments, investigating the effect of purification methods on transfection and/or editing efficiency and repeating the experiments may better resolve whether D839N has a significantly higher efficiency than PE2.

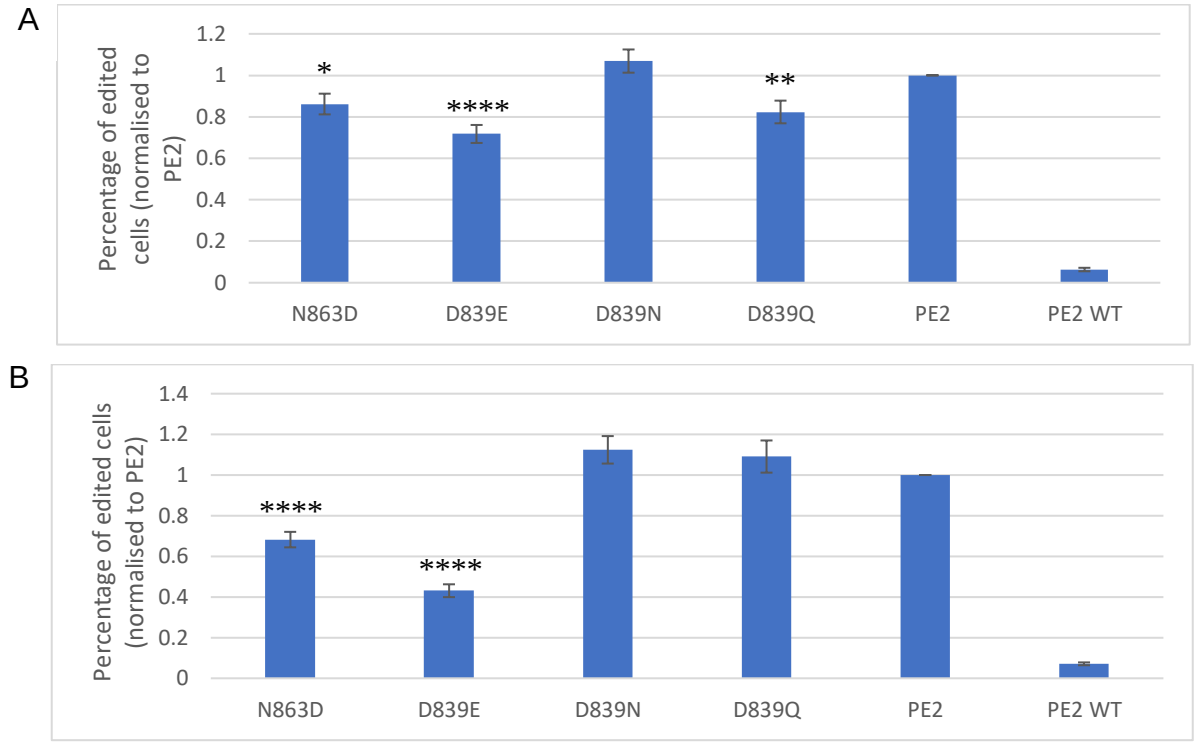

Figure 8. Genome editing efficiencies of PE2 mutants purified with maxiprep

Laser scanning cytometry (acumen Cellista) was used to quantify the percentage of edited cells (%). The more efficient PEG RNA 2/3 (top) and the less efficient PEG RNA 3/2 (bottom) were used respectively. Standard error bars were obtained by dividing standard deviation by the square root of the number of samples (n = 15). The stars show significance as compared to PE2. Significance was determined using a one sample t-test.

A The cells were transfected with the more efficient PEG RNA 2/3 along with the mutated PE2.

B The cells were transfected with the less efficient PEG RNA 3/2 along with the mutated PE2.

Combined PE2 mutants

The mutations resulting in significantly different editing efficiencies (Fig. 7) were combined to find out if that leads to greater editing efficiencies. N863D was combined with D839E, D839N, or D839Q. All combined mutants had significantly lower editing efficiency as compared to PE2 (Fig. 9).

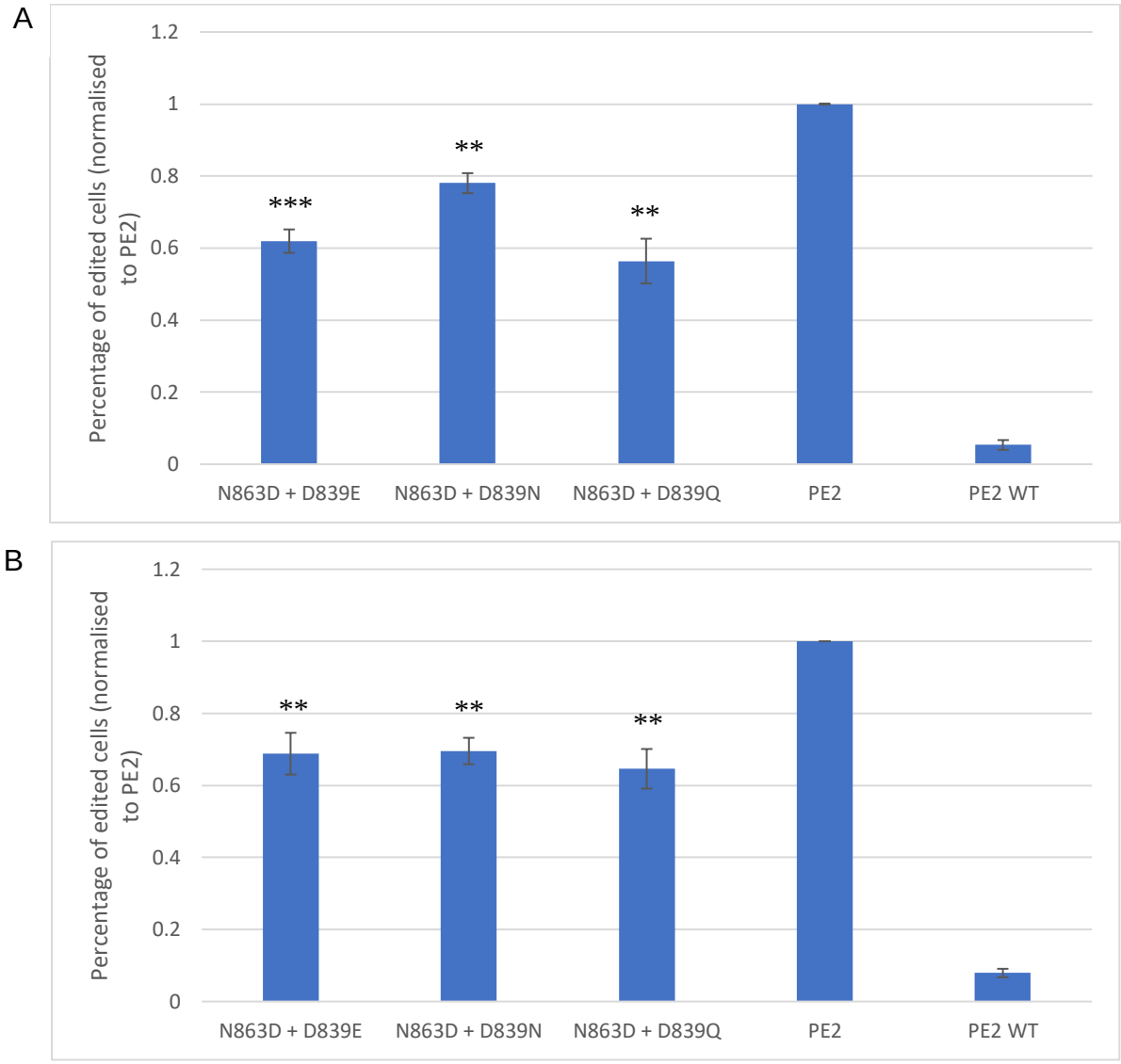

Figure 9. Genome editing efficiencies of combined PE2 mutants purified with miniprep

Laser scanning cytometry (acumen Cellista) was used to quantify the percentage of edited cells (%). The more efficient PEG RNA 2/3 (top) and the less efficient PEG RNA 3/2 (bottom) were used respectively. Standard error bars were obtained by dividing standard deviation by the square root of the number of samples (n = 5). The stars show significance as compared to PE2. Significance was determined using a one sample t-test.

A The cells were transfected with the more efficient PEG RNA 2/3 along with the mutated PE2.

B The cells were transfected with the less efficient PEG RNA 3/2 along with the mutated PE2.

The effects of combining mutations can be understood with a diagram (Fig. 10) based on the hypothesis that slowing the HNH domain can improve the editing efficiency of PE2. The aim is to get to the peak, which represents optimal time-separated cleavage of the DNA strands. Mutations add up to increasingly disabled Cas9. When mutations left of the peak are combined, the resulting editing efficiency can be higher than the individual component mutations. When mutations on both sides of the peak are combined, the editing efficiency could be lower than an individual mutation. Screening for more mutations and combining them would allow better understanding of the mutational landscape.

It should be noted that while the mutations have been designed to slow the HNH domain, to conclusively prove that slowing the HNH domain has the potential to improve PE2, the kinetics of the mutants have to be measured. Additionally, factors other than kinetics may come into play with the combinations having another effect on the enzyme. Although the mutations are in the HNH domain, they can also affect overall stability and folding. This could have decreased editing efficiency further.

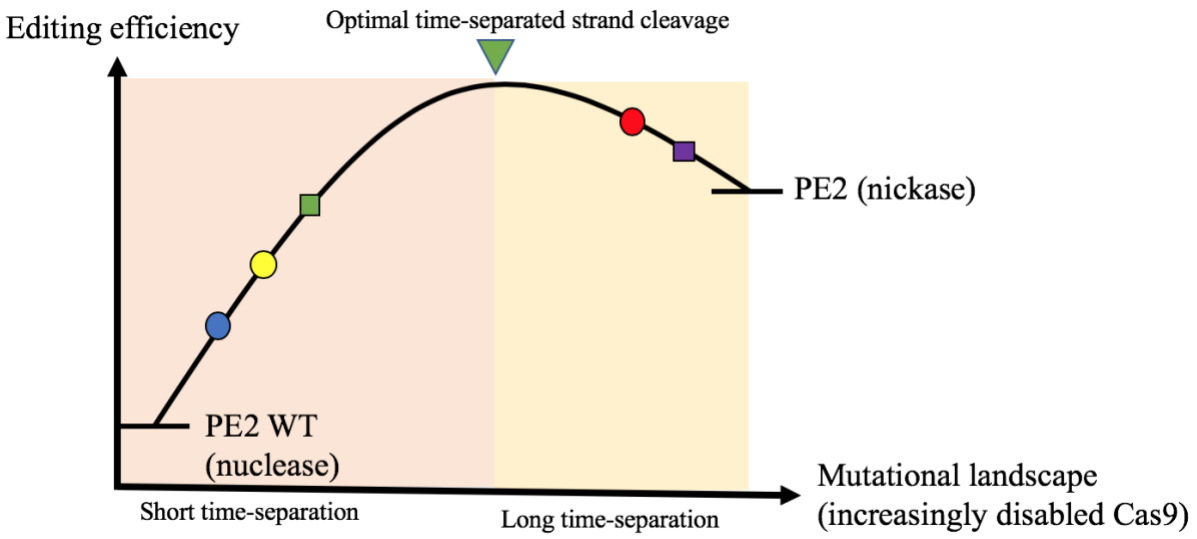

Figure 10. Schematic diagram illustrating how editing efficiencies may vary along mutational landscape to explain editing efficiencies when mutations are combined

Blue and yellow circles represent mutations left of the peak in editing efficiency while the red circle represents a mutation on the right side of the peak. The green square represents combining two mutations on the same side (blue and yellow) resulting in increased efficiency while the purple square represents combining two mutations on different sides (blue and red) resulting in decreased efficiency.

High throughput screening of mutations

Based on the initial promising results, we extended the mutagenesis screen to other sites likely to play a role in catalysis and adapted the protocols to a plate format. Twenty mutations that potentially give rise to lowered editing efficiency were identified (Fig. 11 and Fig. 12): H840K, H840Q, H840N, H840R, D837E, D837N, E779Y, K866R, Q807D, Q807K, Y823E, R905K, S909N, D912Q, D868E, I852R, T804R, T804N, T804E, and S851D. They were selected for showing editing efficiencies different from both PE2 and PE2 WT. The twenty selected mutants were sequenced and shown to carry the correct mutation.

Four mutations identified, H840K, H840Q, H840N and H840R, all involve the histidine in the HNH active site. In a related structure Endo VII, His41 acts as a general base and activates a water molecule for catalysis (Nishimasu et al, 2014). His840 in Cas9 corresponds to His41 in Endo VII. In their deprotonated forms, lysine and arginine can both act as general bases. However, due to their higher pKas of 10.5 and 12.5 for lysine and arginine respectively as compared to 6.0 for histidine, at physiological pH there is a smaller proportion of deprotonated forms of lysine and arginine available to act as bases as compared to histidine. While the deprotonated fraction would be negligible based on the pKa values stated, the pKa values are modified by the surroundings. The smaller proportion of deprotonated forms results in less efficient strand cleavage. For asparagine and glutamine, the lone pair of electrons on nitrogen are partially delocalised into the carbonyl group and hence less available for activating the water molecule. Therefore, these mutations may slow down the general base catalysis involved in bond cleavage. Another possible explanation would be the different positioning of the nitrogen atom.

The mutations D837E and D837N affect Cas9 similarly to the D839N mutation mentioned previously. Both Asp837 and Asp839 are located in the HNH catalytic cleft and are involved in binding a Mg^{2+} ion through their carboxyl groups. This Mg^{2+} ion coordinates with the oxygen atoms of the phosphate group between bases 3 and 4 upstream of the PAM (Huai et al, 2017; Chen, 2014; Jinek et al, 2012; Gasiunas et al, 2012; Biertümpfel et al., 2007; Li et al., 2003). Through a one-metal-ion hydrolysis mechanism, the phosphodiester bond is split. Through substituting aspartic acid for similar amino acids that can still coordinate Mg^{2+} but with altered strength of interaction or position, the hydrolysis may be slowed down.

Cryo-EM data shows that in the "product" complex of Cas9, disorder in loops L1 (residues 766-779) and L2 (residues 907-925) accompany changes in the HNH domain which becomes disordered and REC2 which is ordered and interacts with nucleic acids (Zhu et al, 2019). The mutation E779Y was chosen as it moves in the active state conformation. Molecular dynamics simulation techniques have arrived at a similar conclusion that during conformational change to the active state,

the HNH domain makes new interactions with the REC lobe that when weakened or disrupted, may lead to less efficient strand cleavage (Zuo & Liu 2017). Glu779 is located on the HNH flanking linker L1 and binds to Arg586 and Lys558 of REC3 (Nishimasu et al, 2014). REC3 is a domain in the Cas9 recognition lobe (REC) that interacts with the RNA/DNA heteroduplex. It has been proposed as an allosteric effector which recognises the heteroduplex, changes conformation upon binding to the target, and then allows for the activation of the HNH domain (Zuo & Liu 2017; Jiang et al, 2017; Palermo et al, 2016; Anders et al, 2014; Nishimasu et al, 2014). The E779Y mutation might have affected the formation of the active state conformation and thus kinetics of the HNH domain.

In changing to the active state, the HNH domain also forms interactions with the sgRNA. Gln807, which is part of an N-terminal ββα motif flanking helix, interacts with the eighth nucleotide from the most PAM-distal end of the sgRNA through hydrogen-bonds and/or salt bridges (Zuo & Liu 2017). The mutation Q807D introduces a negative charge and the mutation Q807K introduces a positive charge, both of which affect these interactions.

Y823E has also been selected for having a lowered editing efficiency as compared to PE2. Tyr823 has an important role in stabilising the catalytic Asp839 through hydrogen-bonding. Asp839 is involved in catalysis by contributing a coordination ligand to Mg^{2+} at the catalytic centre. Stabilisation of Asp839 by Tyr823 allows Asp839 to orientate optimally for coordination and catalysis (Chen et al, 2017). Replacing tyrosine with glutamic acid introduces a negative charge which affects the stabilisation of Asp839 and hence results in less efficient strand cleavage.

Another residue that interacts with Asp839 is Lys866. The amino of Lys866 is close enough to hydrogen bond with the side-chain carboxyl of Asp839. It has been proposed that this interaction is important for Mg^{2+} coordination based upon its significance in *Actinomyces naeslundii* and HmuI Cas9 (Zhu et al, 2019).

Mutations R905K, S909N, D912Q are located in the loop formed by residues 906 to 918 which interacts with the non-target strand to favour the approach of the HNH domain towards the target strand cleavage site (Palermo et al, 2018; Palermo et al, 2016; Nishimasu, 2014). Hence, weakening these interactions through the abovementioned mutations destabilise HNH at the cleavage site. This alters the dynamics of its activation and slows its cleavage activity (Chen et al, 2017; Slaymaker et al, 2016).

Other mutations are likely to affect hydrogen bonding interactions between amino acids or affect the arrangement of water molecules around the active site.

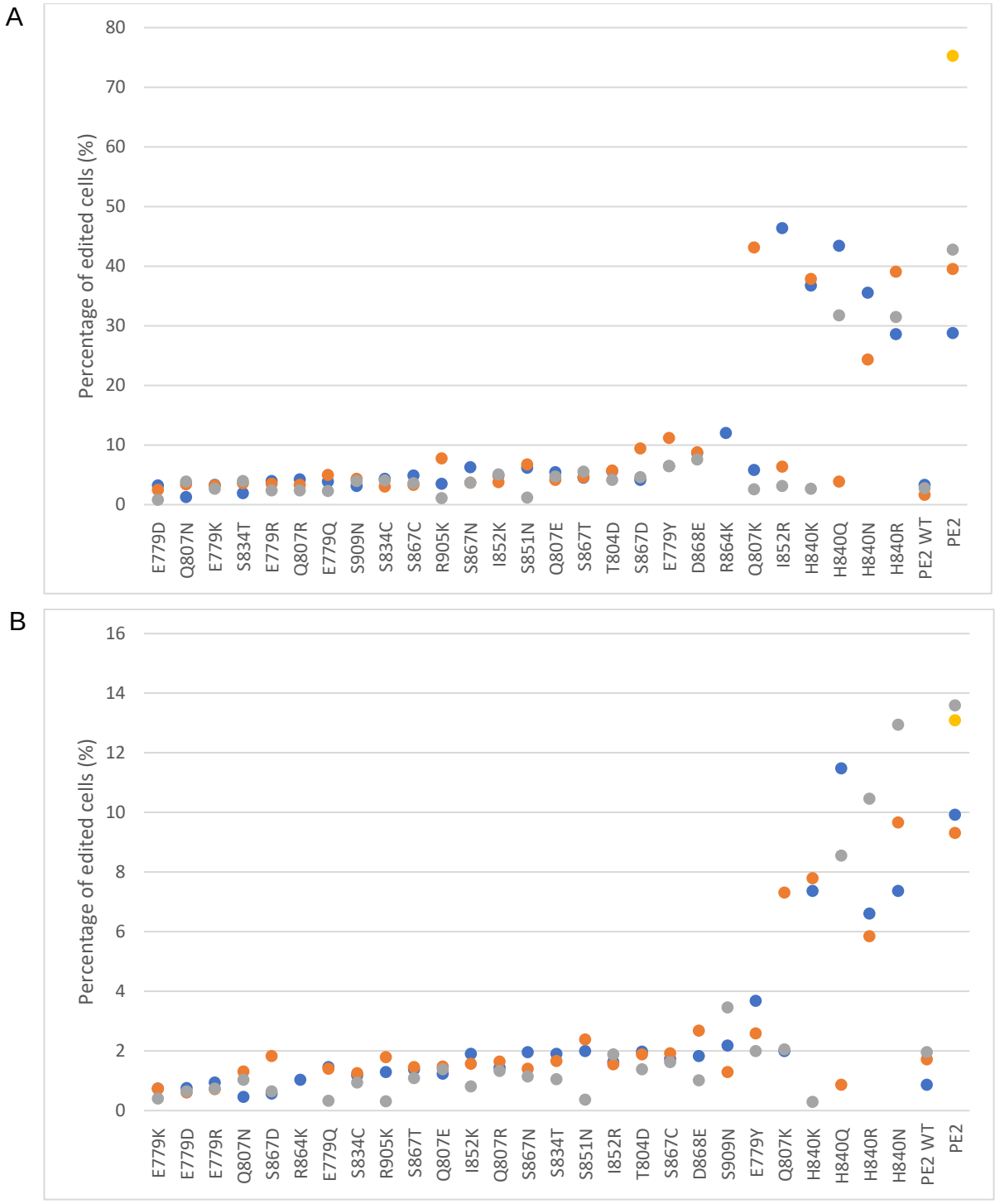

Figure 11. Genome editing efficiencies of PE2 mutants purified with Wizard® SV 9600

Laser scanning cytometry (acumen Cellista) was used to quantify the percentage of edited cells (%). Each point represents the editing efficiency of PE2 purified from liquid culture of a single colony. As mutagenesis may not succeed for all colonies, there will be some points vastly lower than other points of the same mutation. These points may not carry the mutation and should be disregarded.

A The cells were transfected with the more efficient PEG RNA 2/3 along with the mutated PE2.

B The cells were transfected with the less efficient PEG RNA 3/2 along with the mutated PE2.

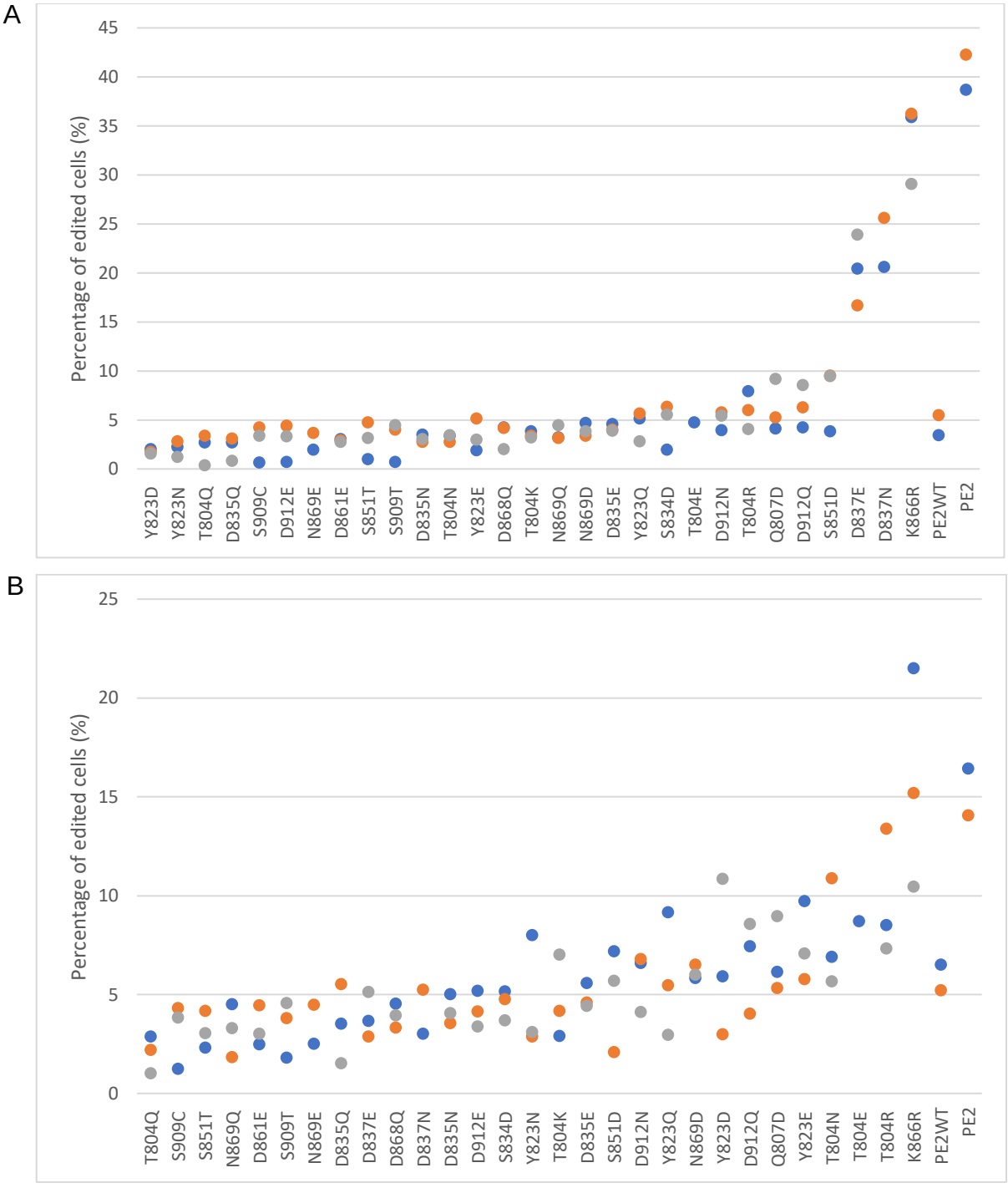

Figure 12. Genome editing efficiencies of PE2 mutants purified with Wizard® SV 9600

Laser scanning cytometry (acumen Cellista) was used to quantify the percentage of edited cells (%). Each point represents the editing efficiency of PE2 purified from liquid culture of a single colony. As mutagenesis may not succeed for all colonies, there will be some points vastly lower than other points of the same mutation. These points may not carry the mutation and should be disregarded.

A The cells were transfected with the more efficient PEG RNA 2/3 along with the mutated PE2.

B The cells were transfected with the less efficient PEG RNA 3/2 along with the mutated PE2.

Conclusion, limitations, and further work

Project limitations include the limited number of mutations screened due to time constraints. Further work can involve screening more mutations and combining them to understand the mutational landscape better. To screen mutants faster, it would help to develop an *in vitro* screen. We have already generated a reporter construct with a point mutation and frameshift mutations in firefly luciferase. Editing by PE2 would revert this mutation. *In vitro* transcription and translation would then show whether the edit was carried out and editing efficiency can be determined. *In vitro* translation would take hours instead of one week which the *in vivo* screen requires. Additionally, the luciferase assay would be more sensitive and work better for distinguishing low efficiency from no efficiency.

Error-prone PCR and directed mutation can generate mutants that cannot be predicted from current structural knowledge. This can be done by screening PE2 mutants from low fidelity PCR using an on-target broken antibiotic marker that has to be cleaved and edited to work and a different off-target antibiotic marker that should not be cleaved or edited to work.

Additionally, to have a more stringent selection process in the *in vivo* screen, the broken GFP plasmid should require multiple rounds of editing to fix. This way, multiple pegRNAs would be required, so the PE2 mutant selected would need to have decent editing efficiency for all pegRNAs instead of only elevating activity when used with a specific pegRNA.

While the mutations were designed to slow down the HNH domain and a mutation has been identified to have increased editing efficiency, it has yet to be conclusively proven that it is this time-separated cleavage of the DNA strands which causes increases in editing efficiency. To demonstrate this mechanism, native gels could be run at certain time points.

Here, we report a rationally engineered Cas9 variant (D839N) that improves editing efficiency in PE2 (Anzalone, 2019). Additionally, we found that a prime editor with both endonuclease domains catalytically active can still edit and can be used in genome editing to introduce large deletions. This can potentially happen by adding complementary sequences at two distant positions. The region between the two sites would be lost. Furthermore, we found 22 other mutations which result in editing efficiency higher than the abovementioned double strand-cutting prime editor but lower than PE2. These mutations can be combined to potentially make a prime editor with improved editing efficiency.

Materials and Methods

Plasmid design and cloning

The PE2 construct pCMV-PE2 (Anzalone et al, 2019) was ordered from Addgene (plasmid 132775). The pegRNAs were expressed in a pU6-based vector (Anzalone et al, 2019) ordered from Addgene (plasmid 132777). Golden gate cloning (New England BioLabs) was used to assemble the pegRNA from 3 pieces (protospacer, constant backbone, and 3' extension with reverse transcription-priming region). The oligonucleotides were ordered from Integrated DNA Technologies and the resultant pegRNAs were purified using the Plasmid Miniprep Kit (Qiagen).

Mutagenesis

Mutations to PE2 were introduced using the Quikchange II Mutagenesis Kit (Agilent) following primer design using the QuikChange® Primer Design Program. Colonies were selected on an Luria-Bertani (LB) agar gel (Formedium) containing 1% carbenicillin. Selected colonies were cultured overnight in LB broth (Gibco) or terrific broth (Gibco). Plasmids were purified using Plasmid Miniprep or Maxiprep Kits (Qiagen) or Wizard® SV 9600 Plasmid DNA Purification System (Promega). Sequences mutated were confirmed by Sanger sequencing (Cambridge Biochemistry DNA Sequencing Facility).

General mammalian cell culture conditions

HEK293T-BFP/GFP cells were cultured and passaged in Dulbecco's modified Eagle's medium (DMEM) plus GlutaMAX (ThermoFisher Scientific) supplemented with 10% FBS and 100 μg mL^{-1} penicillin streptomycin (Gibco). Cells were incubated, maintained, and cultured at 37 °C with 5% CO_2.

HEK293T tissue culture transfection protocol

HEK293T-BFP/GFP cells were seeded on 96-well plates (BD Falcon). 24 hours after seeding, media was changed to DMEM plus GlutaMAX (ThermoFisher Scientific) supplemented with 10% FBS without antibiotics. Cells were transfected with 37 ng of PE2-expressing plasmids and 37 ng of pegRNA-expressing plasmids using a 4:1 ratio of FuGENE® HD transfection reagent volume (μL) to DNA mass (μg) according to manufacturer's protocol. 4 μL of transfection mixture was added to 100 μL of media in each well. 24 hours after transfection, media was changed back to Dulbecco's modified Eagle's medium (DMEM) plus GlutaMAX (ThermoFisher Scientific) supplemented with 10% FBS and 100 μg mL^{-1} penicillin streptomycin (Gibco). 96 hours after transfection, cells were trypsinised, resuspended, and diluted to 25% of the original cell density in preparation for laser scanning cytometry.

Laser Scanning Cytometry

Acumen® Cellista was used to measure the number of cells expressing GFP and/or BFP. Excitation wavelengths were 488 nm for GFP and 405 nm for BFP and emission collection windows were 500-530 for GFP and 420-490 for BFP. Editing efficiency was calculated by dividing the number of green objects by the number of blue objects.

Acknowledgements

I would like to thank Dr. Otto Kauko, my day-to-day supervisor, for all his guidance throughout this project. I would also like to thank Prof. Jussi Taipale, my overall supervisor, for his mentorship. Special thanks goes to Dr. Minna Taipale for her help and support and Dr. Ekaterina Morgunova for advising about mutation designs, as well as all the other members of the Taipale lab for their kind assistance.

References

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

Anzalone A, Randolph P, Davis J, Souza A, Koblan L, Levy J, Chen P, Wilson C, Newby G, Raguram A & Liu D (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576: 149–157

Bhaya D, Davison M & Barrangou R (2011) CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics* **45:** 273–297

Biertumpfel C, Yang W & Suck D (2007) Crystal structure of T4 Endonuclease VII N62D mutant in complex with a DNA Holliday junction. *Nature* **449:** 616–620

Chan K, Tong AHY, Brown KR, Mero P & Moffat J (2019) Pooled CRISPR-Based Genetic Screens in Mammalian Cells. *Journal of Visualized Experiments* **151:** e59780

Chapman JR, Taylor MR & Boulton SJ (2012) Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell* **47:** 497–510

Chen H, Choi J & Bailey S (2014) Cut Site Selection by the Two Nuclease Domains of the Cas9 RNA-guided Endonuclease. *Journal of Biological Chemistry* **289:** 13284–13294

Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A & Doudna JA (2017) Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550:** 407–410

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA & Zhang F (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339:** 819–823

Cotten, M, Baker A, Saltik M, Wagner E and Buschle M (1994) Lipopolysaccharide is a frequent contaminant of plasmid DNA preparations and can be toxic to primary human cells in the presence of adenovirus. Gene Therapy **1:** 239-246

Cox DBT, Platt RJ & Zhang F (2015) Therapeutic genome editing: prospects and challenges. *Nature Medicine* **21:** 121–131

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J & Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471:** 602–607

Gasiunas G, Barrangou R, Horvath P & Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109:** 2579–2586

Gottesman S Microbiology: Dicing defence in bacteria. *Nature* **471:** 588–589

Haapaniemi E, Botla S, Persson J, Schmierer B & Taipale J (2018) CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine* **24:** 927–930

Huai C, Li G, Yao R, Zhang Y, Cao M, Kong L, Jia C, Yuan H, Chen H, Lu D & Huang Q (2017) Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nature Communications* **8:**

Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, Kommineni S, Chen J, Sondey M, Ye C, Randhawa R, Kulkarni T, Yang Z, Mcallister G, Russ C, Reece-Hoyes J, Forrester W, Hoffman GR, Dolmetsch R & Kaykas A (2018) p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine* **24:** 939–946

Jiang F, Zhou K, Ma L, Gressel S & Doudna JA (2015) A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348:** 1477–1481

Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E & Doudna JA (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351:** 867–871

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337:** 816–821

Li CL, Hor LI, Chang ZF, Tsai LC, Yang WZ & Yuan HS (2003) DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. *The EMBO Journal* **22:** 4014–4025

Mali P, Yang L, Esvelt KM, Aach J, Guell M, Dicarlo JE, Norville JE & Church GM (2013) RNA-Guided Human Genome Engineering via Cas9. *Science* **339:** 823–826

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F & Nureki O (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156: 935–949

Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA & McCammon JA (2018) Key role of the REC lobe during CRISPR-Cas9 activation by "sensing", "regulating" and "locking" the catalytic HNH domain. Q Rev Biophys. 51: 91

Palermo G, Miao Y, Walker RC, Jinek M & Mccammon JA (2016) Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Central Science* **2:** 756–763

Rouet P, Smih F & Jasin M (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* **91:** 6064–6068

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX & Zhang F (2015) Rationally engineered Cas9 nucleases with improved specificity. *Science* **351:** 84–88

Terns MP & Terns RM (2011) CRISPR-based adaptive immune systems. *Current Opinion in Microbiology* **14:** 321–327

Wang T, Larcher L, Ma L & Veedu R (2018) Systematic Screening of Commonly Used Commercial Transfection Reagents towards Efficient Transfection of Single-Stranded Oligonucleotides. *Molecules* **23:** 2564

Watson JF & García-Nafría J (2019) In vivo DNA assembly using common laboratory bacteria: A re-emerging tool to simplify molecular cloning. Journal of Biological Chemistry 294: 15271–15281

Weber M, Möller K, Welzcek M & Schorr J (1995) Short technical reports. Effects of lipopolysaccharide on transfection efficiency in eukaryotic cells. Biotechniques 19: 931-939

Wiedenheft B, Sternberg SH & Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482:** 331–338

Zhu X, Clarke R, Puppala AK, Chittori S, Merk A, Merrill BJ, Simonović M & Subramaniam S (2019) Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nature Structural & Molecular Biology* **26:** 679–685

Zuo Z & Liu J (2017) Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Scientific Reports* **7:** 17271