Premature termination codons signaled targeted gene repair by nonsense mRNA-mediated gene editing in *E. coli*

Xiaolong Wang[*1], *Haibo Peng*[1], *Chunyan Li*[1], *Xuxiang Wang*[1], *Yalei Wang, Gang Chen*[1], *Jianye Zhang*[1]

College of Life Sciences, Ocean University of China, Qingdao, 266003, P. R. China

[1] To whom correspondence should be addressed: *Xiaolong Wang, Ph.D., Department of Biotechnology, Ocean University of China, No. 5 Yushan Road, Qingdao, 266003, Shandong, P. R. China*, Tel: *0086-139-6969-3150*, E-mail: *Xiaolong@ouc.edu.cn*.

Abbreviations used in this article: *E. coli: Escherichia coli*; BLA: β-lactamase; *bla*: BLA gene; HSC: hidden stop codon; PTC: premature termination codon; OE-PCR: overlapping extension polymerase chain reaction; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis; TCB: tetracycline-containing broth; TCP: tetracycline-containing plate; ACP: ampicillin-containing plate; ACB: ampicillin-containing broth; KCP: kanamycin-containing plate; KCB: kanamycin-containing broth; 2nd NGS: the second-generation high-throughput sequencing; DEG: differentially expressed gene; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOBAS: KEGG Orthology Based Annotation System; FPKM: expected number of Fragments PerKilobase of transcript sequence per Millions base pairs sequenced; SNP: Single Nucleotide Polymorphism; InDel: Insertions/Deletion; SV: Structure Variation. DSB: double strand break;

Abstract

Frameshift mutations yield truncated proteins, leading to loss-of-function, genetic disorders or even death. Reverse mutations, which restore the wild-type phenotype of a mutant, were assumed to be far rarer than forward mutations. However, in this study, screening tests showed that the revertants of a frameshift mutation were detected more frequently than expected in *E. coli*. Sanger sequencing revealed that reverse mutations were caused not by random mutagenesis but by active targeted gene repair. Molecular studies suggested that it was the premature termination codons (PTCs) in nonsense mRNAs that signaled the repair of the frameshift mutation. Genome survey indicated that the genome sequence of a revertant is not more variable than that of a wild-type strain. Transcriptome profiling identified differentially expressed genes and pathways that were upregulated in frameshift or revertant which possibly involved in frameshift repair, include DNA replication, RNA surveillance, RNA editing, mismatch repair and homologous recombination. Introducing synthetic DNA or RNA oligonucleotides into the mutant increased the recovery rates as they promoted the frameshift repair. Based on these data, we hypothesized a molecular model for frameshift repair referred to as *nonsense mRNA-mediated gene editing* (NMGE): nonsense mRNAs were recognized by mRNA surveillance by PTCs signaling, edited by RNA editing and used to direct the repair of their defective coding gene through mismatch repair and homologous recombination. In addition, this mechanism also serve as a driving force for molecular evolution, and the widespread presence of frameshift homologs within and across species is considered as evolutionary evidences preserved in nature.

1. Introduction

DNA replication and DNA repair happens daily in cell division and multiplication. DNA damages, point mutations and InDels can be induced by chemical and physical mutagens, such as radiation, pollutants or toxins. In addition, because of the imperfect nature of DNA replication, mutations and InDels also occur spontaneously by replication errors or slipped strand mispairing.

If the length of an InDel is not a multiple-of-three, it causes a frameshift mutation in a protein-coding gene, results in a substantial change of its encoded amino acids in the downstream; in addition, it also often produces a number of nonsense codons and yields truncated and non-functional proteins [1], which may lead to a genetic disorder or even death [2-5]. The transcripts from a frameshifted coding gene are recognized as nonsense mRNAs by mRNA surveillance [6-12], and the nonsense codons emerged in a frameshifted gene or mRNA are called HSCs [1], also known as PTCs [13, 14].

The molecular mechanisms/pathways involved in the repair of DNA damages and point mutations have been intensively studied and well-understood [15], including Mismatch Repair, Base Excision Repair, Nucleotide Excision Repair, Homologous Recombination and Non-Homologous End Joining. The maintenance of the reading frames of coding genes is fundamentally important for cells. However, the molecular mechanism for the repair of frameshift mutation (frameshift repair) remains unknown. The reverse mutation phenomenon was discovered as early in the 1960s [16]. Forward and reverse mutation have been both explained by random mutagenesis, and it has been assumed that reverse mutations occur at a far lesser rate than forward mutations. However, here we report that reverse mutations occur more frequently than expected in *E. coli*, and it is not a random mutagenesis but an active targeted gene repair.

2. Materials and Methods

2.1 Frameshift mutagenesis and back mutation screening

A plasmid containing a frameshifted β-lactamase gene (*bla*), pBR322-(bla-), was constructed by site-directed mutagenesis (Fig 1): in plasmid pBR322, a G:C base pair located in the upstream (+136) of the wild-type *bla* gene (*bla*+) was deleted by

OE-PCR. Competent cells of *Escherichia coli* strain DH5α was transformed with pBR322-(bla-), propagated in TCB, and the transformed bacteria were plated on a TCP to count the total number of bacteria and plated on ACPs to screen for revertants. The screened revertants were picked and propagated in ACB at 37°C with 200 rpm shaking and with 1 mL of overnight seed culture in 10 mL ACB, their plasmids DNA were extracted and their *bla* genes were Sanger sequenced. The recovery rate was calculated by the number of revertants divided by the total number of bacteria. The growth rates were evaluated by the duration to reach the later log phase.

2.2 Construction and expression of a PTC-free frameshifted bla

A *PTC-free frame-shifted bla* (denoted as *bla#*) was constructed by an *in vitro readthrough* technique: (1) the *bla#* sequence was derived from *bla-* by replacing each nonsense codon with a sense codon according to the *in vivo readthrough rules* (Table 1), and adding a stop codon (TAA) in the 3'-end; (2) the *bla#* gene was chemically synthesized by *Sangon Biotech, Co. Ltd* (*Shanghai*), inserted into an expression vector, pET28a, and transformed into competent cells of *E. coli* strain BL21. The transformed bacteria were plated on a KCP, the grown transformants were picked and propagated in KCB, plated on ACPs to screen for revertants. The revertants were subcultured in ACB to test their growth rates and their *bla* sequences were Sanger sequenced; (3) The expression of the frameshifted BLA was induced by IPTG. Total protein samples were extracted, frameshifted BLA was purified by nickel column chromatography and analyzed by SDS-PAGE. The activity of BLA was measured by iodometry.

2.3 Introduction of DNA or RNA oligonucleotides into the frameshift mutant

Competent cells of the frameshift mutant, which contains the defective plasmid pBR322-(bla-), were transformed with a synthetic DNA or RNA oligonucleotides with a 39-nt wild-type sequence restoring the base pair which was deleted in OE-PCR. The transformed cells were cultured in TCB at 37°C for 6 hours with 200 rpm shaking, and then plated on TCPs to count the total number of bacteria, and on ACPs to screen for the induced revertants, whose *bla* gene were repaired. The induced revertants were subcultured in ACB to test their growth rates, and their *bla* were Sanger sequenced. The recovery rate was calculated by the number of revertants divided by the total

number of bacteria. The growth rates were evaluated by the duration to reach the later log phase by culturing at 37°C with 200 rpm shaking and with 1 mL of overnight seed culture in 10 mL ACB.

2.4 Genomes resequencing and variation analysis of the E. coli strains

Genomic DNA samples were extracted from wild-type (*bla*+) and revertant (*bla**) strains. Library preparation and genome sequencing were conducted by a commercial service provided by *Novogene Co. Ltd*. The library was sequenced on an Illumina HiSeq 250PE platform and paired-end reads were obtained. Raw reads in fastq format were processed by removing the adapter sequence and low quality reads. The clean reads with high quality were mapped onto the reference genome of *E. coli* K12 MG1655 (NC_000913.3) to identify SNPs, InDels and SVs of each sample. GATK2 software was used to perform SNP calling. Samtools v0.1.18 was used to sort the reads and reorder the bam alignments. Circos was used to display the coverage of the reads and the distribution of SNPs and InDels on the ring diagram. BreakDancer was used to detect structure variations.

2.5 Transcriptome analysis of the E. coli strains

Total RNA samples were extracted from the wild type (*bla*+), the frameshift (*bla*-) and the revertants (*bla**) *E. coli* strains. Library preparation and RNA sequencing were conducted by a commercial service provided by *Novogene Co. Ltd*. After library preparation and cluster generation, the library was sequenced on an Illumina HiSeq platform. The paired-end raw reads were processed by removing the adapter sequence and low quality reads. The clean reads with high quality were mapped onto the annotated reference genome, *E. coli* K12 MG1655 (NC_000913.3), and the expression levels of each gene were calculated for each sample and compared to each other to identify DEGs, enriched GO terms and KEGG pathways.

(1) *Quantification of gene expression level*

Bowtie (v2-2.2.3) was used for index building and aligning the clean reads to the reference genome. HTSeq (v0.6.1) was used to count the numbers of reads mapped to each gene. And then FPKM of each gene was calculated based on their lengths and reads counts.

(2) *Differential expression analysis*

Prior to differential expression gene analysis, for each strain the read counts were adjusted by the edgeR program through one scaling normalized factor. Differential expression analysis of two conditions was performed using the DEGSeq R package (1.20.0). The P values were adjusted by the Benjamini & Hochberg method. Corrected P-value of 0.005 and log2 (Fold change) of 1 were set as the threshold for significantly differential expression.

(3) *GO and KEGG enrichment analysis of differentially expressed genes*

GO enrichment analysis was implemented by the GOseq R package, in which gene length bias was corrected. GO terms with corrected P-value less than 0.05 were considered significantly enriched by DEGs. KOBAS 2.0 was used to test the statistical enrichment of DEGs in KEGG pathways (http://www.genome.jp/kegg/).

3. Results and Analysis

3.1 The growth of the frameshift mutant and the revertants

In the plasmid pBR322 containing a wild-type *bla* gene, one single G:C base pair in the upstream of *bla* was deleted by OE-PCR (Fig 1). In the resulting frameshifted *bla* (*bla*-), seventeen nonsense codons appeared; the active sites of BLA, including its substrate binding site, locate all in the downstream of the deletion. The plasmid pBR322 containing wild-type or frameshifted *bla* gene was transformed into *E. coli* (strain DH5α and BL21). The wild-type bacterial (*bla*+) grew well on both ACPs and TCPs (Fig 2A, left), while it was expected that the frameshift mutant (*bla*-) could not grow in ACPs as the frameshift mutation is a loss-of-function. However, repeatedly a few ampicillin-resistant colonies (Fig 2A, middle) were observed in the ACPs. The possibility of carryover- or cross-contamination was ruled out by blank controls (Fig 2A, right). So, these ampicillin-resistant colonies must be revertants (*bla**) whose *bla*-gene was repaired by reverse mutations.

In addition, although most of the revertants can grow in ACB, the growth rates of different revertants varied greatly and are often much slower when compared to the wild-type strain: by culturing at 37°C with 200 rpm shaking, it took the wild type 12

to 24 hrs. to reach the late log phase, while it took the revertants up to 36 to 48 hrs. to grow, and many of the revertants failed to grow in ACB.

Hitherto, it seems that there is nothing unusual, since reverse mutations are very common phenomenon. Forward and reverse mutation have been both explained by random mutagenesis (Fig 3A): in DNA replication, natural or induced mutations occur randomly, a forward mutation cause a defective mutant or a loss-of-function, and a reverse mutation restore the wild-type phenotype. In the above tests, the revertants survived only because they were 'lucky': their *bla*- gene coincidently restored by a reverse mutation, while most of the other bacteria died without a reverse mutation.

3.2 A reverse mutation is not a random mutagenesis but a targeted gene repair

The random mutagenesis model for reverse mutation sounds faultless. However, in a 'thought experiment', we noticed that the model is indeed inconsistent: (1) in the model, a bacterium survived if its defective *bla* gene was restored through random mutagenesis. However, a reverse mutation must occur in a process of DNA replication or repair in a living cell. Since the frameshift mutant itself could not live in ampicillin, a reverse mutation must have already occurred in a cell before the adding of ampicillin. (2) In *E. coli*, forward mutations occur at a very low rate (~10^{-6}~10^{-8} per locus per generation). It was assumed that reverse mutations occur at a far lesser rate than forward mutations. The reverse mutation rate for a certain frameshift mutation should be orders of magnitudes lower than the forward mutation rate if it depends on random mutagenesis. (3) The bacterial genome consists of millions of base pairs, and thousands of genes. If the cells did not 'know' which gene was defective, there would be little chance to repair it through random mutagenesis. Therefore, reverse mutation should be far lesser than a forward mutation. However, in fact the revertants can always be detected by a routine screening, so reversions occur much more frequently than expected. The average recovery rate measured in DH5α is 2.61×10^{-8}, which is comparable to the baseline forward mutation rate; in BL21, however, it reaches up to 1.19×10^{-7}, which is ~10-fold higher than the baseline.

Therefore, it is rather suspectable that biology relies on random mutagenesis, a simple but uncertain strategy, to address this important issue. It is more likely that a

specific mechanism has been developed for frameshift repair in the history of life. In summary, the reversion of a frameshift mutation is not like a random mutagenesis but more like an active and targeted frameshift repair. However, how is a frameshifted gene being identified and repaired? So far, it is still unknown at the molecular level.

3.3 Sanger sequencing of the bla genes of the revertants

To further investigate how a frameshift gene was repaired, the revertants were picked and propagated in ACB. Their plasmids DNA were extracted and sequenced by the Sanger method. Unexpectedly, in the initial revertants, the *bla* sequence read automatically from their Sanger sequencing diagrams are often neither a wild type nor a reverse mutation, but a frameshift (Fig 4A, top). These revertants were subcultured in ACB, some survived and some not. The initial revertants usually grow very slowly, but the survived subcultures grow faster and faster.

In the later subcultures, the *bla* sequences read automatically from their Sanger sequencing diagrams are often still a frameshift, but their sequencing diagrams often contain two (or three, or more) sets of superposed peaks (Fig 4A, middle), indicating that they consist of two (or three or more) types of *bla* sequences, one is the original frameshifted and the other ones are the repaired. Finally, in some of the fast-growing subcultures, most of their *bla* copies were repaired (Fig 4A, bottom).

By visual inspection, the repaired *bla* sequences of the revertants were read from the secondary peaks in the Sanger diagrams. By aligning the *bla** sequences against the *bla*+ and *bla*- sequence, it was indicated that they contain a variety of InDels and substitutions, especially G:C→A:T transitions (Fig 4B, top), and their encoded amino acid sequences are different from the sequence of the wild-type BLA (Fig 4B, bottom). Moreover, they often still contain some stop codons. In other words, the frameshifted gene was not restored completely but partially and progressively, resulting in a variety of frameshifted variants called *frameshift homologs*.

The tetracycline-resistant gene (*tcr*) in the same plasmid was also sequenced by Sanger method, while no obvious overlapping peak was observed in their sequencing diagrams. Therefore, the high-level variants observed in *bla* is not a result of random mutagenesis, but more like a gene repair targeting specifically to the *bla* gene.

3.4 The revertant is not a mutator as their genome is not highly variable

A mutator strain, such as *E. coli* strain mutD5, has an extraordinary high rate of spontaneous mutagenesis over their whole genome due to its defective proofreading and mismatch repair during DNA replication [17]. To clarify further whether the *bla* variants are the results of whole-genome random mutagenesis, the genome sequences of a frameshift (*bla*+) and a revertant (*bla**) were sequenced by 2nd NGS sequencing, mapped onto the reference genome sequence of the *E. coli* strain K12 and compared with each other. As shown in Fig 4C-4D, Table 2-3, the SNP/InDel levels of the tested revertant is not higher but even lower than that of the frameshift, suggesting that the genome of the revertant was not more variable than that of the frameshift mutant. In addition, the SNP/InDel levels of the frameshift and the revertant are both very low, suggesting that their proofreading/mismatch repair system was not defective and they adopted a stringent strategy in DNA replication. Therefore, the high level of variants observed in *bla* in the revertants is not a whole-genome mutagenesis, but an active gene repair targeting specifically to the *bla* gene.

In addition, genome structures analysis shows that there are quite some structure variations (SVs) in both two genomes tested (Fig 4E). However, most of the SVs size in 100~200 bp, which equals approximately to the length of the reads, therefore, they were considered as falsely-mapped reads rather than true structure variations. As well known, the 2nd NGS platforms can have several limitations, such as short read-length and amplification bias, posing serious problems in genome assembly, mapping and structures analysis [18]. However, thanks to the extremely high depth of sequencing of the 2nd NGS, the SNP/InDel data is considered valid and reliable.

3.5 PTCs signaled the repair of the frameshift mutation

We speculated that it was the PTCs presented in nonsense mRNAs that signaled the repair of the frameshift mutation. To validate this idea, a *PTC-free bla-* (denoted as *bla#*) was derived from *bla-* by replacing each nonsense codon with an appropriate sense codon in accordance with the *readthrough rules* (Table 1). The *bla#* gene was synthesized chemically, cloned in the plasmid vector pET-28a and expressed in *E. coli* strain BL21. The *bla#* gene of the transformants were Sanger sequenced, while few

variations were found. The *bla#* transformants were plated on ACPs, but no revertant was observed. A ~34-kDa band was detected by SDS-PAGE (Fig 5A), representing the product expressed from the PTC-free frameshifted *bla*. No lactamase activity was detected by iodimetry test of the product (the frameshifted BLA) (Fig 5B), suggesting that it is indeed a loss-of-function.

In summary, the *bla#* gene was kept unchanged and a dysfunctional product was expressed in its wrong reading frame. The only difference between *bla-* and *bla#* is: the former gene consists of a certain number of nonsense codons, but the latter one does not; therefore, it must be the PTCs appeared in the transcripts (nonsense mRNAs) that signaled the repair of the frameshift mutation in *bla-*.

3.6 Identifying genes/proteins involved in frameshift repair

We speculated that frameshift repair requires the upregulation of relevant genes, proteins and pathways. By transcriptomes sequencing and profiling, 292 or 247 genes displayed upregulated transcriptions (Supplementary dataset 1-2), 65 or 58 pathways were identified as upregulated in the pathway enrichment (Supplementary dataset 3-4), respectively in the frameshift (*FS*) or revertant (*RE*) when compared to the wild-type (*W*). As shown in Fig 6B-6C and Table 4-5, upregulated pathways that are probably involved in frameshift repair were manually selected, including *DNA replication* (Fig 6D), *RNA surveillance/processing/degradation* (Fig 6E), *DNA mismatch repair* and *homologous recombination* (Fig 6F).

In addition, by carefully inspecting the list of upregulated genes, we noticed that the cytosine or isoguanine deaminase gene (*codA*) was significantly upregulated in the frameshift. CodA catalyzes the hydrolytic deamination of cytosine to uracil. CodA also catalyzes deamination of isoguanine, a mutagenic oxidation product of adenine in DNA, and of isocytosine. As mentioned above, Sanger sequencing suggested that the repaired *bla* contains many G:C→A:T substitutions, suggested that they were derived from deamination of DNA bases. It has been reported that deamination of DNA bases generates deoxyinosine from deoxyadenosine and links mismatch repair and RNA editing [19]. Here we found that the deaminase and the mismatch repair pathway were both upregulated in the frameshift, suggesting that RNA editing and mismatch repair

were probably both involved in the frameshift repair. However, further experimental investigations are needed to validate the regulation and the function of these genes in the repair of frameshift mutations.

3.7 A new model for frameshift repair: nonsense mRNA-mediated gene editing

At present, the mechanism for the recognition and repair of frameshift mutation is unknown at the molecular level. However, we postulated that the transcribed nonsense mRNAs must be involved, because PTCs cannot function in a double helix DNA but can only in a translating nonsense mRNA.

We conducted an in-depth survey of previous studies that covered a wide range of prokaryotic and eukaryotic organisms. We put these fragmentary evidences together by presuming that the underlying mechanism of frameshift repair is highly conserved among prokaryotic and eukaryotic species. Based on our studies and others, here we hypothesized a molecular model for frameshift repair, *nonsense mRNA-mediated gene editing* (NMGE), consisting five main steps (Fig 3B):

(1). ***Nonsense mRNA recognition***: when the frameshifted gene was transcribed and translated, produced nonsense mRNAs containing some PTCs. When a nonsense mRNA is translated, the ribosome is blocked when it encounters a PTC [20] , which were identified by mRNA surveillance through PTCs signaling.

(2). ***Nonsense mRNA processing***: most of the nonsense mRNAs were degraded by *nonsense-mediated mRNA decay* (NMD). In both prokaryotes and eukaryotes, NMD is the best-studied nonsense mRNA processing pathways [21-23]. NMD targeting mRNAs harboring PTCs for degradation enhances their decay rate and reduces the abundance of the mRNA transcribed from that gene. In bacterial cells, mRNA quality is also tightly controlled but it is not yet clear how the nonsense mRNAs are processed [24]. In eukaryotes, it has been reported that NMD uses the presence of an exon junction complex (EJC) in the downstream of a nonsense codon as a second signal to distinguish a PTC from a true termination codon [25]. Factors that regulate NMD include three interacting proteins, UPF1, UPF2 (also known as NMD2) and UPF3, which were encoded by highly conserved genes originally identified in yeast [26, 27]. In addition to NMD, nonsense mRNAs are

also subject to other mRNA processing pathways, including *nonsense-mediated translational repression* (NMTR), *nonsense-mediated transcriptional silencing* (NMTS), and *nonsense-mediated alternative splicing* (NMAS).

(3). ***Nonsense mRNA translation*:** If the target gene is essential, a cell survives if and only if it can produce a functional protein by translating the nonsense mRNAs. It has been reported that frameshifted or nonsense coding genes can be expressed through some special mechanisms, including *translational frameshifting* [28-35] and *translational readthrough* [36-39]. Translational frameshifting, also known as *programmed ribosomal frameshifting*, is a process occurs in all known organisms, from *E. coli* to mammals [40-43], whereby the ribosome is guided toward a triplet nucleotide that is shifted either one nucleotide position upstream (+1) or one nucleotide position downstream (−1). In yeast, telomere maintenance is controlled by ribosomal frameshifting and the NMD pathway [35]. Translational readthrough is a process wherein the nonsense mRNAs can be translated by translating each PTC into an appropriate amino acid [44-48]. Mutations in *upf* genes caused nonsense suppression and subsequent translational readthrough in yeast and 5~20 fold decreases in the rates of NMD [49], simultaneously, suggested a direct link between NMD and nonsense suppression. The three Upf proteins all interact with Sup35, and Upf1 interact with Sup45 (eRF1), a second release factor, and binding of a SMG-1-Upf1-eRF1-eRF3 (SURF) complex to EJC triggers Upf1 phosphorylation and NMD [50]. In addition, the nonsense suppression caused by *upf* mutations is additive to those by *sup35* and *sup45* mutations [51], suggesting that nonsense suppression is a nonsense mRNA processing pathway alternative to NMD. In other words, when NMD is inhibited, nonsense mRNAs can be translated by nonsense suppression, translational frameshifting or translational readthrough.

(4). ***RNA editing***: as described in the above, nonsense mRNAs must be involved in the recognition and repair of the frameshift mutation. Since a nonsense mRNA by itself is defective and it cannot template the gene repair directly, we assumed that the nonsense mRNAs were functionalized by editing prior to templating the gene

repair. By transcription, each copy of the target gene produces many copies of nonsense mRNAs, and thus, the probability of producing a functional protein by mRNA editing is significantly better than that of random mutagenesis in its coding DNA. RNA editing has been studied extensively and intensively [52-56], and the mechanisms of RNA editing, such as insertion/deletion of uridine [57], has been reported in kinetoplastid, trypanosome [58] and human [59], suggesting that the molecular mechanisms for RNA editing are widely exists and highly conserved. In particular, the *apoB* mRNA editing complex suppresses NMD [60], suggesting that sometimes nonsense mRNAs are subjected to RNA editing instead of NMD. As described above, in the revertants, distinct types of bases were substituted, inserted or deleted in the *bla* gene. These editing events are located mostly near to, or in the downstream, but not in the distant upstream of the base pair deleted in the mutagenesis (Fig 4A-4B). This phenomenon can be explained by the PTCs in the nonsense mRNAs: the PTCs presented in the downstream of the deletion not only signal the recognition of nonsense mRNA but serve as flags for RNA editing. Therefore, the frameshifted gene might be repaired through editing the nonsense mRNAs rather than repair the coding sequence directly. A cell can recognize the nonsense mRNAs by PTC signaling, but it might be impossible to locate the exact site of the frameshift-causing InDel, so it is not always possible to restore the original coding sequence. Therefore, the cell must insert, delete and/or substitute nucleotides randomly within or nearby the PTCs, resulting in a variety of variants (frameshift homologs). In a word, by PTC-signaling, mRNA editing is restricted near to or in the downstream of the InDel in the nonsense mRNAs, so that DNA editing will also be restricted in the appropriate regions in the target coding sequence.

(5). ***RNA-directed DNA repair***: After editing, the translation of the nonsense mRNA restarted. If existed, the remaining PTCs were readthrough by translating each of them into an amino acid. A cell survived when it obtained a functional protein by editing and translation of the nonsense mRNAs. The edited mRNAs were then transported back into the nucleus or plasmid DNA to localize and direct the repair

of its own defective coding gene. Although this process is unknown in details, the mismatch repair and the homologous recombination pathways must be involved, as they were both upregulated in the frameshift and the revertants (Fig 6F).

3.8 *The NMGE model better explains the growth of the revertants*

As mentioned above, the frameshift repair is not a sudden event but a progressive process, the growth rate of a revertant changes and those of different revertants vary. Interestingly, the sizes of the wild-type colonies are uniform (Fig 2), showing that their growth rates are uniform; while the sizes of the revertant colonies varied greatly (Fig 7A), indicating that their growth rates differed due to the activities of their BLA were diversified. Comparing with the random mutagenesis, NMGE better explains the diversified growth rates of the revertants and also the high-level variations observed in their *bla* genes, but not in the other genes in their genome or the plasmid:

(1). In the initial revertants, their *bla* genes were still frameshifted, their survival can be explained in three possible ways: (a) the frameshifted *bla* is translated by *ribosomal frameshifting*, produced a wild-type BLA; (b) the frameshifted *bla* is translated by *translational readthrough*, produced a frameshifted but functional BLA; or (c) the initial slowly-growing revertants may also contain a few copies of repaired *bla*, but the proportion of the repaired *bla* was too low to be detected by Sanger sequencing. As pBR322 is a multi-copy plasmid, it might be sufficient to support the survival of the host cell in ACM even when only a small proportion of its copies were repaired. Because these mechanisms are all very time- and resource-consuming, so that the host bacteria grow very slowly, and many of them died if their *bla* was not repaired soon.

(2). By NMGE, their *bla* were repaired progressively, not only before but also after the adding of ampicillin. Ampicillin is the substrate of BLA, it is neither the cause nor the terminator of the gene repair, but a selection pressure to screen and display the revertants (the result of repair).

(3). NMGE is signaled by the PTCs presented in the *bla* nonsense mRNAs, and the aim of NMGE is to eliminate the PTCs. In the frameshifts, there were many PTCs in the *bla* mRNAs, the probability of producing a functional product by

mRNA editing is very low, so that their growth rate is very slow and their survival rate is very low, as most of them died before their *bla* were repaired.

(4). In the latter revertants, their *bla* were repaired by NMGE, with more and more nonsense codons were replaced by sense codons, fewer and fewer stop codons remained, and thus, gene repairing become easier and easier and their growth rates become faster and faster.

(5). Finally, gene repair completed when all nonsense codons were eliminated. The *bla* coding sequence was recovered, not necessarily the same as the wild-type, but consist of a variety of variants.

Previously, we reported that frameshift homologs are widespread within and across species, as the genetic code was optimized to tolerate frameshifting [61]. This study further demonstrated that it is the PTCs that signaled the frameshift repair. Frameshift mutations occurred in a protein-coding gene are repaired through NMGE, resulting in a variety of frameshift homologs. Because of the shiftability of the coding genes, many of these frameshift homologs are functional and was preserved in nature. Therefore, the widespread existence of frameshift homologs is evolutionary evidences for this model. In short, NMGE is not only a mechanism for frameshift repair but also a driving force for molecular evolution.

3.9 Both DNA and RNA efficiently direct the repair of frameshift mutation

Synthetic DNA oligonucleotides or DNA/RNA complexes have been widely used for targeted gene repair [62-64]. And it has also been reported that synthetic RNA or endogenous transcripts can direct the repair of double-strand breaks of DNA [65-70]. We postulated that a synthetic wild-type RNA could direct the frameshift repair if the NMGE model is correct. To validate this, competent cells of the frameshift mutants (*E. coli* strain DH5α or BL21) were transformed by heat-shock at 42 °C with a 39-nt synthetic sense (ss-), antisense (as-) or double-stranded (ds-) DNA or RNA oligonucleotides with a wild-type (WT) or mutated (MT) sequence to recover the base which was deleted in the OE-PCR. The recovery rate (f_r) was very low and unstable in DH5α, but much higher and more stable in BL21. We performed statistics analysis on the data of BL21. As shown in Fig 7A-7B, in the control or heat-shock groups, f_r was

1.19~1.30×10^{-7}. When an ss-, as- or ds-DNA oligo was added, f_r increased compared with the control groups. f_r increased 1~2 fold in ss-DNA group, 3~6 fold in as-DNA group and 2~4 fold in ds-DNA group. These data are consistent with previous studies reported that both as- and ss-DNA can induce DNA-directed targeted gene repair [71]. Wild-type (WT) ss-RNA caused 0.5-fold decrease of f_r. Surprisingly, however, the mutated (MT) ss-RNA promoted the gene repair and caused 2~5-fold increase of f_r in BL21, which is consistent with our expectation. No or few revertants were observed in asRNA group, probably due to asRNA can silence the transcripts and result in the death of the host.

For each group, the *bla* genes were sequenced to test whether they were repaired or not. The Sanger diagram showed that a designed base was inserted into the repaired *bla* gene in both the ssDNA (Fig 7C) and the ssRNA group (Fig 7D). In the ssDNA or ssRNA group, the type of the insertion base is mostly consistent with the designed type, and the insertion was located exactly at the site of deletion; in the control group, however, the type and location of the inserted base are both unpredictable, suggesting that, like DNA, RNA by itself can also direct targeted gene repair.

Homologous recombination must play an important role in frameshift repair, as the average recovery rate measured in BL21 (recA+) is significantly higher than that of DH5α (recA-). However, the DNA- or RNA-induced recovery rates were increased both in BL21 and DH5α, suggesting that homologous recombination is not the only pathway responsible for frameshift repair. In all DNA, RNA and control groups, G:C→A:T substitutions were often observed in the repaired *bla* (Fig 7B); in addition, in the ssRNA group, not only G:C→A:T substitutions were often observed, but also a ssRNA designed with an G→A substitution efficiently introduced the substitution into the repaired *bla* (Fig 7C), resulted in an increased recovery rate, which is even higher than that of the wild-type ssRNA (Fig 7B), suggesting that mismatch repair and/or RNA editing might be involved in the repair.

4. Discussion

4.1 The mechanism for the repair of frameshift mutations

The molecular mechanisms for the repair of DNA damage and point mutations have been well studied [15], however, so far no mechanism specifically designed for the repair of frameshift mutations has been discovered. This NMGE mechanism not only well explains our observations but is supported by many previous studies.

When a frameshift mutation occurs in a protein-coding gene in a cell, it does not repair the defective gene directly, but first edit the transcripts and then use an edited mRNA to direct the frameshift repair by NMGE. This is very surprising, but in fact it is more reasonable than direct gene editing: a frameshifted coding gene must be first transcribed before it can be recognized by RNA surveillance, because in principle a frameshift mutation cannot be recognized at the genomic-DNA level.

To repair a frameshifted gene by NMGE, as shown by a question mark in Fig 3B, there exist two possible strategies: (1) ***direct NMGE***: use an original nonsense mRNA to identify the frameshifted coding gene, and then edit the defective gene by DNA-level mutagenesis; however, the edited coding gene must be transcribed and translated for a second round to check whether it is repaired or not; (2) ***indirect NMGE***: the nonsense mRNAs are first edited, translated by translational readthrough, transported to identify the coding gene, and then used to template gene repair. *i.e.*, a nonsense mRNA was functionalized through RNA editing prior to directing the repair of their own coding gene. Obviously, compared to the direct strategy, indirect NMGE might be more reasonable and more efficient, since the repair happens in a living cell only when it has obtained a functional mRNA by RNA editing.

4.2 Evidences for the connection of DNA repair and RNA editing

All types of RNAs are subject to either degradation or processing and a variety of molecular/cellular mechanisms are involved in RNA processing [72]. In the last few decades, studies have established a link between DNA repair and RNA processing. For example, several proteins that respond to DNA damage and base excision repair enzymes, such as SMUG1, APE1 and PARP1, have been shown to participate in RNA surveillance and processing [73].

Deamination is the removal of an amine group from a molecule. Enzymes that catalyze this reaction are called deaminases. Deamination of DNA bases generates

deoxyinosine from deoxyadenosine, create missense mutations predisposing humans to cancer and interfere with other basic molecular genetic processes. Cytidine deamination was discovered as a mechanism of editing mRNA transcripts first in eukaryotes [74, 75]. The recent identification of tRNA-specific adenosine deaminases (ADATs) has led to the suggestion that these enzymes, as well as the cytidine and adenosine deaminases acting on pre-mRNAs (CDARs and ADARs), belong to a superfamily of RNA-dependent deaminases [76]. The original and best-characterized example of C→U editing is the apolipoprotein B (apoB) mRNA, in which a single, site-specific cytidine deamination introduces a stop codon (UAA) into the reading frame of the edited transcript, leading to production of the shortened isoform, apoB48. This occurs exclusively in the human small intestine, where apoB48 is required for the absorption of dietary lipid [77]. At present, it has been well known that RNA editing by base deamination plays many important roles and aberrant or dysregulated RNA editing might contribute to genomic instability in cancer [78].

In addition, several studies have provided evidences that deamination is related to both DNA repair and RNA editing. For example, endonuclease V, which is highly conserved from *E. coli* to human [19], is also involved in excision repair that removes deoxyinosine from DNA. It is reported that human endonuclease V, localized to the cytoplasm, is also a ribonuclease specific for inosine-containing RNA, hydrolyses the second phosphodiester bond located in the 3' to the inosine in unpaired regions in the double-stranded RNA, and controls the fate of inosine-containing RNA in humans [19]. In addition, editing of the pre-mRNA for DNA repair enzyme NEIL1 causes a lysine to arginine change in the lesion recognition loop of the NEIL1 protein, and the recoding site is a preferred site for the RNA editing adenosine deaminase ADAR1 [79]. Moreover, it has been reported that engineered fusions of CRISPR/Cas9 and a cytidine deaminase enzyme mediated the conversion of cytidine to uridine, effecting a C→T or G→A substitution [80]. These studies all suggested a direct link between the regulation of mismatch repair and RNA editing. In the above, we demonstrated that the repaired *bla* in the revertants often contains many C→T (or G→A) substitutions, and the cytidine deaminase and the mismatch repair genes were both upregulated in

the frameshift mutant and revertant tested. We therefore speculate that both mismatch repair and RNA editing are involved in frameshift repair.

4.3 NMGE depends on RNA-directed DNA Repair

As is well known, RNA is transcribed in the nucleus and then transported to the cytoplasm [81]. Recently, there are growing evidences suggesting that RNAs can be transported back to the nucleus to repair DNA double strand breaks (DSBs), a process known as *RNA-directed DNA Repair*. Not only a synthetic DNA/RNA hybrid, but also a RNA-only oligonucleotide can serve as a template for DNA synthesis for the repair of a DSB in a homologous DNA. It was reported that short RNA patches can direct modifying DNA in *E. coli* [82, 83], yeast [69, 70] and human embryonic kidney (HEK-293) cells [83]. To our knowledge, however, this is the first report that, RNA by itself can direct precise and targeted repair of a frameshift mutation. At present, we are further investigating on how RNA guides the targeted gene repair.

NMGE depends on, but is not equivalent to, RNA-directed DNA Repair, since it is triggered by PTCs and the nonsense mRNAs are edited prior to directing the repair of the coding DNA. As mentioned in the above, there are many evidences suggest a direct or close link between RNA editing and DNA repair, which can support this NMGE model. However, further systematic investigations are needed to elucidate this process in details and validate it in other species.

4.4 The possibility of RNA-only gene and genome editing

CRISPR/Cas9 for RNA-guided genome editing [84-86] has been widely used for the introduction or correction of specific mutations [87, 88], which can potentially be used in gene therapy to repair a disease-causing frameshift/point mutation. However, RNA-guided genome editing requires the transfection of foreign genes in the host cell. On the other hand, deoxyoligonucleotides-based targeted gene repair (ODN-TGR) [62, 63] could be more favorable in human and animal due to its non-transgenic properties. Unfortunately, however, it suffered from inefficiency and uncertainty because it relies on the induction of the endogenous DNA repair system while it is suppressed by the mismatch repair mechanisms [64].

In addition, synthetic DNA oligonucleotides or DNA/RNA complexes have been widely used for targeted gene repair [62-64]. And it has also been reported that synthetic RNA or endogenous transcripts can direct the repair of double-strand breaks of DNA [65-70], but it has never been reported that endogenous or synthetic RNAs can be edited and used to direct targeted gene repair. The essential components of NMGE, including mismatch repair, homologous recombination and RNA editing, are highly conserved from bacteria to human, so potentially genes and genomes can be edited using RNA-only molecules without introducing any exogenous gene or protein.

Author Contributions

Xiaolong Wang conceived the study, designed the experiments, analyzed the data and wrote the paper; Chunyan Li, Haibo Peng, Xuxiang Wang and Yalei Wang performed the experiments; Gang Chen and Jianye Zhang provided materials, equipment, give suggestions and discussed on the paper.

Acknowledgements:

This study was supported by the *National Natural Science Foundation of China* (Grant No. 81072567). We appreciate Dr. Quanjiang Dong from the Department of Gastroenterology, Qingdao Municipal Hospital for helpful suggestions and discussions.

Figure legends

Fig 1. The introduction of a frameshift mutation in the upstream of *bla* in the plasmid pBR322. Sanger sequencing diagram of: (A) the wild type (*bla*+); (B) the frameshift mutant (*bla*-). (C) Alignment of the nucleotide sequence of t *bla*+ and *bla*-.

Fig 2. Growth of *E. coli* on ACPs and TCPs: *bla*+: Wild-type; *bla*-: Frameshift; *bla**: revertants; *blank*: Blank control

Fig 3. Two different models for the repair of frameshift mutations:

(A) The traditional "random mutagenesis" model: **Step 1**: when a frameshift mutation occurred in a coding gene, a certain number of hidden stop codons (HSCs) (*red bars*) emerged in the CDS, and in the nonsense mRNAs, caused *translational termination* and resulted in truncated products; **Step 2**: when the CDS was repaired by a *reverse mutation* due to random mutagenesis, the CDS was restored (fully or partially), the stop codons were hidden (*green bars*), and the translation proceed, producing functional mRNAs and proteins;

(B) This nonsense-mediated gene editing (NMGE) model: **Step 1:** when the frameshifted gene was transcribed and translated, the nonsense mRNAs containing some PTCs (*yellow bars*) were identified by mRNA surveillance through PTCs signaling; **Step 2**: most nonsense mRNAs caused *translational termination* were degraded by *nonsense-mediated mRNA decaying* (NMD); **Step 3**:

some nonsense mRNAs were translated by ribosomal frameshifting or translational readthrough; **Step 4**: some nonsense mRNAs were repaired by *RNA editing*; **Step 5**: nonsense or repaired mRNAs were transported to identify and repair the defective coding gene, the stop codons were hidden (*green bars*), the CDS restored partially and the translation proceed, the remaining stop codons were readthrough; **Step 6:** the coding gene was repaired partially by *RNA-directed DNA repair*, producing functional mRNAs and proteins, and so that the host recovered gradually.

Fig 4. Sanger Sequencing and sequence analysis of the *bla* gene of the revertants:

(A) The Sanger diagram of bla: **Top**: in the initial revertants, most *bla* was still frameshifted; **Middle**: in the later subcultures, sequencing diagram contains two sets of overlapped peaks, one is frameshifted *bla*, the other is repaired; **Bottom**: in the final subcultures, most of their *bla* were repaired;

(B) Alignment of the wild-type, the frameshifted and the repaired bla coding sequence (top) and their translated proteins sequence (bottom): *bla* coding sequence were read artificially from the secondary peaks of the Sanger diagram of the revertants and their subcultures. *bla*: wild-type; *bla-*: frameshift; *A-E*: different revertants; *red box*: the base deleted in the OE-PCR; *blue boxes*: bases inserted or deleted in the revertants;

(C) Circos view of the genome sequences of the revertant and the frameshift;

(D) The distribution of SNPs/InDels on the genome of wild-type and revertants;

(E) The reported "structure variations (SVs)" in the genome of wild-type and revertants;

Fig 5. Expression of the PTC-free frameshifted *bla* (*bla#*) in pET28a-*bla#* in *E. coli* BL21:

(A) SDS-PAGE detection of the frameshifted BLA; Lane 1: uninduced; lane 2: induced by 1M IPTG; lane 3: uninduced superficial ultrasound crushing; lane 4: induced by ultrasonic breakdown of the supernatant; lane 5: the ultrasonic crushing of the uninduced precipitate; lane 6: the ultrasonic crushing of the precipitate induced;

(B) Detection of β-lactamase activity by iodimetry: (1) *E. coli* BL21/pET28a-bla- caused fading, suggesting that their *bla-* was repaired, produced active β-lactamase, transformed ampicillin into penicillium thiazole acid, which compete with starch for iodine; (2) *E. coli* BL21/pET28a-bla# expressed a frameshifted product (BLA#), but did not cause fading, suggesting that the product is not active, and that *bla#* was not repaired; (3) The negative control (*E. coli* BL21), no β-lactamase, no fading.

Fig 6. The transcriptome analysis of different *E. coli* strains: *W_1*: wildtype; *FS_2*: frameshift; *R_3*: initial revertant; *RE_4*: subculture of a revertant; (A) The number of differential expression genes (DEGs) and the cluster analysis of DEGs; (B) the most enriched GO terms by comparing the frameshift (FS_2) with the wild type (W_1); (C) the most enriched GO terms by comparing the revertant (R_3) with the frameshift (FS_2); (D) Genes upregulated in the RNA degradation/processing pathway; (E) Genes upregulated in the mismatch repair and homologous recombination pathway.

Fig 7. Targeted repair of the frameshift mutation induced by DNA or RNA oligonucleotides.

(A) Both ssDNA and ssRNA oligonucleotides can induce the repair of the frameshift mutation; (B)

The spontaneous or induced reversion rates in different group; (C) Sanger sequencing diagram of the DNA-induced revertants; (D) Sanger sequencing diagram of the RNA-induced revertants;

Table 1. The natural suppressor tRNA for nonsense mutations (the *readthrough rules*).

Site	tRNA (AA)	Wild type		Correction	
		Code	**Anti-code**	**Code**	**Anti-code**
supD	Ser (S)	→ UCG	CGA←	→ UAG	CUA←
supE	Gln (Q)	→ CAG	CUG←	→ UAG	CUA←
supF	Tyr (Y)	→ UAC	GUA←	→ UAG	CUA←
supG	Lys (K)	→ AAA	UUU←	→ UAA	UUA←
supU	Trp (W)	→ UGG	CCA←	→ UGA	UCA←

Table 2. The summary of InDel in the *bla**/- genome

Sample	Insertion	Deletion	Het	Hom	HetRate(%)	Total
bla-	2	4	0	6	0	6
*bla**	2	3	0	5	0	5

Table 3. The summary of SNP in the *bla*+/ - genome

Sample	ts	tv	ts/tv	Het	Hom	HetRate(%)	Total	Density(SNP/Kb)
bla-	43	27	1.59	4	66	0	70	0.02
*bla**	42	27	1.56	2	67	0	69	0.01

1 Table 4. Possible NMGE-relevant genes and pathways that were upregulated in the frameshift

Pathway Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
ABC transporters	KEGG	eco02010	32	171	0.0001	0.0090
RNA degradation	KEGG	eco03018	1	15	0.7419	0.9762
DNA replication	KEGG	eco03030	1	17	0.7822	0.9762
Mismatch repair	KEGG	eco03430	1	22	0.8575	0.9796
Homologous recombination	KEGG	eco03440	1	27	0.9069	0.9796

2

3

4 Table 5. Possible NMGE-relevant genes and pathways that were upregulated in the revertant

Pathway Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
ABC transporters	KEGG	eco02010	32	171	3.53E-05	0.0018
Homologous recombination	KEGG	eco03440	2	27	0.6539	0.9519
RNA degradation	KEGG	eco03018	3	15	0.1017	0.5470
DNA replication	KEGG	eco03030	1	17	0.7564	0.9519
Mismatch repair	KEGG	eco03430	1	22	0.8357	0.9519

5

6

References

1. Seligmann, H. and D.D. Pollock, *The ambush hypothesis: hidden stop codons prevent off-frame gene reading.* DNA Cell Biol, 2004. **23**(10): p. 701-5.
2. Owens, M., et al., *SOS1 frameshift mutations cause pure mucosal neuroma syndrome, a clinical phenotype distinct from multiple endocrine neoplasia type 2B.* Clin Endocrinol (Oxf), 2016. **84**(5): p. 715-9.
3. Yan, S.E., et al., *In-depth analysis of hyaline fibromatosis syndrome frameshift mutations at the same site reveal the necessity of personalized therapy.* Hum Mutat, 2013. **34**(7): p. 1005-17.
4. Amiel, J., et al., *Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome.* Nat Genet, 2003. **33**(4): p. 459-61.
5. Kang, S., et al., *GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome.* Nat Genet, 1997. **15**(3): p. 266-8.
6. Ge, Z., et al., *Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway.* Elife, 2016. **5**.
7. Schweingruber, C., et al., *Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells.* Biochim Biophys Acta, 2013. **1829**(6-7): p. 612-23.
8. Niu, D.K. and J.L. Cao, *Nucleosome deposition and DNA methylation may participate in the recognition of premature termination codon in nonsense-mediated mRNA decay.* FEBS Lett, 2010. **584**(16): p. 3509-12.
9. Muhlemann, O., et al., *Recognition and elimination of nonsense mRNA.* Biochim Biophys Acta, 2008. **1779**(9): p. 538-49.
10. Muhlemann, O., *Recognition of nonsense mRNA: towards a unified model.* Biochem Soc Trans, 2008. **36**(Pt 3): p. 497-501.
11. Muhlrad, D. and R. Parker, *Recognition of yeast mRNAs as "nonsense containing" leads to both inhibition of mRNA translation and mRNA degradation: implications for the control of mRNA decapping.* Mol Biol Cell, 1999. **10**(11): p. 3971-8.
12. Zhang, J. and L.E. Maquat, *Evidence that the decay of nucleus-associated nonsense mRNA for human triosephosphate isomerase involves nonsense codon recognition after splicing.* RNA, 1996. **2**(3): p. 235-43.
13. Fujikura, K., *Premature termination codons in modern human genomes.* Sci Rep, 2016. **6**: p. 22468.
14. Wong, T.Y., et al., *Role of premature stop codons in bacterial evolution.* J Bacteriol, 2008. **190**(20): p. 6718-25.
15. Jensen, N.M., et al., *An update on targeted gene repair in mammalian cells: methods and mechanisms.* J Biomed Sci, 2011. **18**: p. 10.
16. Streisinger, G., et al., *Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday.* Cold Spring Harb Symp Quant Biol, 1966. **31**: p. 77-84.
17. Schaaper, R.M., *Mechanisms of mutagenesis in the Escherichia coli mutator mutD5: role of DNA mismatch repair.* Proc Natl Acad Sci U S A, 1988. **85**(21): p. 8126-30.

18. Roberts, R.J., M.O. Carneiro, and M.C. Schatz, *The advantages of SMRT sequencing.* Genome Biol, 2013. **14**(7): p. 405.

19. Kuraoka, I., *Diversity of Endonuclease V: From DNA Repair to RNA Editing.* Biomolecules, 2015. **5**(4): p. 2194-206.

20. Hwang, J. and Y.K. Kim, *When a ribosome encounters a premature termination codon.* BMB Rep, 2013. **46**(1): p. 9-16.

21. Le Hir, H., et al., *The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay.* EMBO J, 2001. **20**(17): p. 4987-97.

22. Gonzalez, C.I., W. Wang, and S.W. Peltz, *Nonsense-mediated mRNA decay in Saccharomyces cerevisiae: a quality control mechanism that degrades transcripts harboring premature termination codons.* Cold Spring Harb Symp Quant Biol, 2001. **66**: p. 321-8.

23. Zhang, J. and L.E. Maquat, *Evidence that translation reinitiation abrogates nonsense-mediated mRNA decay in mammalian cells.* EMBO J, 1997. **16**(4): p. 826-33.

24. Richards, J., et al., *Quality control of bacterial mRNA decoding and decay.* Biochim Biophys Acta, 2008. **1779**(9): p. 574-82.

25. Scofield, D.G., X. Hong, and M. Lynch, *Position of the final intron in full-length transcripts: determined by NMD?* Mol Biol Evol, 2007. **24**(4): p. 896-9.

26. He, F., A.H. Brown, and A. Jacobson, *Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway.* Mol Cell Biol, 1997. **17**(3): p. 1580-94.

27. Leeds, P., et al., *The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon.* Genes Dev, 1991. **5**(12A): p. 2303-14.

28. Xu, J., R.W. Hendrix, and R.L. Duda, *Conserved translational frameshift in dsDNA bacteriophage tail assembly genes.* Molecular Cell, 2004. **16**(1): p. 11-21.

29. Baykal, U., A.L. Moyne, and S. Tuzun, *A frameshift in the coding region of a novel tomato class I basic chitinase gene makes it a pseudogene with a functional wound-responsive promoter.* Gene, 2006. **376**(1): p. 37-46.

30. Chen, J., et al., *Dynamic pathways of -1 translational frameshifting.* Nature, 2014. **512**(7514): p. 328-32.

31. Gaudin, C., et al., *Structure of the RNA signal essential for translational frameshifting in HIV-1.* J Mol Biol, 2005. **349**(5): p. 1024-35.

32. Bekaert, M., et al., *Identification of programmed translational -1 frameshifting sites in the genome of Saccharomyces cerevisiae.* Genome Res, 2005. **15**(10): p. 1411-20.

33. Licznar, P., et al., *Programmed translational -1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs.* EMBO J, 2003. **22**(18): p. 4770-8.

34. Klobutcher, L.A. and P.J. Farabaugh, *Shifty ciliates: frequent programmed translational frameshifting in euplotids.* Cell, 2002. **111**(6): p. 763-6.

35. Advani, V.M., A.T. Belew, and J.D. Dinman, *Yeast telomere maintenance is globally controlled by programmed ribosomal frameshifting and the nonsense-mediated mRNA decay pathway.* Translation (Austin), 2013. **1**(1): p. e24418.

36. Claisse, M.L., M. Boguta, and W. Zagorski, *Translational readthrough of a termination codon in the yeast mitochondrial mRNA VAR1 as a result of mutation in the release factor mRF1.* Acta Biochim Pol, 2005. **52**(1): p. 129-37.

37. Halvey, P.J., D.C. Liebler, and R.J. Slebos, *A Reporter System for Translational Readthrough of Stop Codons in Human Cells.* FEBS Open Bio, 2012. **2**: p. 56-59.

38. Eswarappa, S.M., et al., *Programmed translational readthrough generates antiangiogenic VEGF-Ax.* Cell, 2014. **157**(7): p. 1605-18.

39. Staff, P.G., *Correction: Functional Translational Readthrough: A Systems Biology Perspective.* PLoS Genet, 2016. **12**(11): p. e1006434.

40. Namy, O., et al., *A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting.* Nature, 2006. **441**(7090): p. 244-7.

41. Dinman, J.D., *Control of gene expression by translational recoding.* Adv Protein Chem Struct Biol, 2012. **86**: p. 129-49.

42. Dinman, J.D., *Mechanisms and implications of programmed translational frameshifting.* Wiley Interdiscip Rev RNA, 2012. **3**(5): p. 661-73.

43. Ketteler, R., *On programmed ribosomal frameshifting: the alternative proteomes.* Front Genet, 2012. **3**: p. 242.

44. Kopelowitz, J., et al., *Influence of codon context on UGA suppression and readthrough.* J Mol Biol, 1992. **225**(2): p. 261-9.

45. Weiss, R.B., *Ribosomal frameshifting, jumping and readthrough.* Curr Opin Cell Biol, 1991. **3**(6): p. 1051-5.

46. Engelberg-Kulka, H., L. Dekel, and M. Israeli-Reches, *Regulation of the escherichia coli tryptophan operon by readthrough of UGA termination codons.* Biochem Biophys Res Commun, 1981. **98**(4): p. 1008-15.

47. Seidman, J.S., B.D. Janssen, and C.S. Hayes, *Alternative fates of paused ribosomes during translation termination.* J Biol Chem, 2011. **286**(36): p. 31105-12.

48. Su, D., Y. Li, and V.N. Gladyshev, *Selenocysteine insertion directed by the 3'-UTR SECIS element in Escherichia coli.* Nucleic Acids Res, 2005. **33**(8): p. 2486-92.

49. Leeds, P., et al., *Gene products that promote mRNA turnover in Saccharomyces cerevisiae.* Mol Cell Biol, 1992. **12**(5): p. 2165-77.

50. Kashima, I., et al., *Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay.* Genes Dev, 2006. **20**(3): p. 355-67.

51. Keeling, K.M., et al., *Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in S. cerevisiae.* RNA, 2004. **10**(4): p. 691-703.

52. Kugita, M., et al., *RNA editing in hornwort chloroplasts makes more than half the genes functional.* Nucleic Acids Res, 2003. **31**(9): p. 2417-23.

53. Bock, R., *Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing.* Biochimie, 2000. **82**(6-7): p. 549-57.

54. Horvath, A., E.A. Berry, and D.A. Maslov, *Translation of the edited mRNA for cytochrome b in trypanosome mitochondria.* Science, 2000. **287**(5458): p. 1639-40.

55. Yoshinaga, K., et al., *Extensive RNA editing and possible double-stranded structures determining editing sites in the atpB transcripts of hornwort chloroplasts.* Nucleic Acids Res, 1997. **25**(23): p. 4830-4.

56. Sloof, P., et al., *RNA editing in mitochondria of cultured trypanosomatids: translatable mRNAs for NADH-dehydrogenase subunits are missing.* J Bioenerg Biomembr, 1994. **26**(2): p. 193-203.

57. Simpson, L., S. Sbicego, and R. Aphasizhev, *Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business.* Rna, 2003. **9**(3): p. 265-76.

58. Alfonzo, J.D., O. Thiemann, and L. Simpson, *The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria.* Nucleic Acids Res, 1997. **25**(19): p. 3751-9.

59. Zougman, A., et al., *Evidence for insertional RNA editing in humans.* Curr Biol, 2008. **18**(22): p. 1760-5.

60. Chester, A., et al., *The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay.* EMBO J, 2003. **22**(15): p. 3971-82.

61. Wang, X.W., X.; Chen, G.; Zhang, J.; Liu, Y.; Yang C. , *The shiftability of protein coding genes: the genetic code was optimized for frameshift tolerating.* PeerJ PrePrints 2015. **3** p. e806v1.

62. de Semir, D. and J.M. Aran, *Targeted gene repair: the ups and downs of a promising gene therapy approach.* Curr Gene Ther, 2006. **6**(4): p. 481-504.

63. Parekh-Olmedo, H., et al., *Gene therapy progress and prospects: targeted gene repair.* Gene Ther, 2005. **12**(8): p. 639-46.

64. Dekker, M., C. Brouwers, and H. te Riele, *Targeted gene modification in mismatch-repair-deficient embryonic stem cells by single-stranded DNA oligonucleotides.* Nucleic Acids Res, 2003. **31**(6): p. e27.

65. Trott, D.A. and A.C. Porter, *Hypothesis: transcript-templated repair of DNA double-strand breaks.* Bioessays, 2006. **28**(1): p. 78-83.

66. Keskin, H., C. Meers, and F. Storici, *Transcript RNA supports precise repair of its own DNA gene.* RNA Biol, 2016. **13**(2): p. 157-65.

67. Keskin, H., et al., *Transcript-RNA-templated DNA recombination and repair.* Nature, 2014. **515**(7527): p. 436-9.

68. Wei, L., et al., *DNA damage during the G0/G1 phase triggers RNA-templated, Cockayne syndrome B-dependent homologous recombination.* Proc Natl Acad Sci U S A, 2015. **112**(27): p. E3495-504.

69. Storici, F., *RNA-mediated DNA modifications and RNA-templated DNA repair.* Curr Opin Mol Ther, 2008. **10**(3): p. 224-30.

70. Storici, F., et al., *RNA-templated DNA repair.* Nature, 2007. **447**(7142): p. 338-41.

71. Liu, L., et al., *Strand bias in targeted gene repair is influenced by transcriptional activity.* Mol Cell Biol, 2002. **22**(11): p. 3852-63.

72. Isken, O. and L.E. Maquat, *Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function.* Genes Dev, 2007. **21**(15): p. 1833-56.

73. Vohhodina, J., D.P. Harkin, and K.I. Savage, *Dual roles of DNA repair enzymes in RNA biology/post-transcriptional control.* Wiley Interdiscip Rev RNA, 2016.

74. Yu, W. and W. Schuster, *Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria.* J Biol Chem, 1995. **270**(31): p. 18227-33.

75. Blanc, V., S. Litvak, and A. Araya, *RNA editing in wheat mitochondria proceeds by a deamination mechanism.* FEBS Lett, 1995. **373**(1): p. 56-60.

76. Gerber, A.P. and W. Keller, *RNA editing by base deamination: more enzymes, more targets, new mysteries.* Trends Biochem Sci, 2001. **26**(6): p. 376-84.

77. Young, S.G., *Recent progress in understanding apolipoprotein B.* Circulation, 1990. **82**(5): p. 1574-94.

78. Anant, S. and N.O. Davidson, *Hydrolytic nucleoside and nucleotide deamination, and genetic instability: a possible link between RNA-editing enzymes and cancer?* Trends Mol Med, 2003. **9**(4): p. 147-52.

79. Yeo, J., et al., *RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1.* Proc Natl Acad Sci U S A, 2010. **107**(48): p. 20715-9.

80. Komor, A.C., et al., *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.* Nature, 2016. **533**(7603): p. 420-4.

81. Nakielny, S. and G. Dreyfuss, *Transport of proteins and RNAs in and out of the nucleus.* Cell, 1999. **99**(7): p. 677-90.

82. Thaler, D.S., G. Tombline, and K. Zahn, *Short-patch reverse transcription in Escherichia coli.* Genetics, 1995. **140**(3): p. 909-15.

83. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutat Res, 2011. **717**(1-2): p. 91-8.

84. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-7.

85. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.

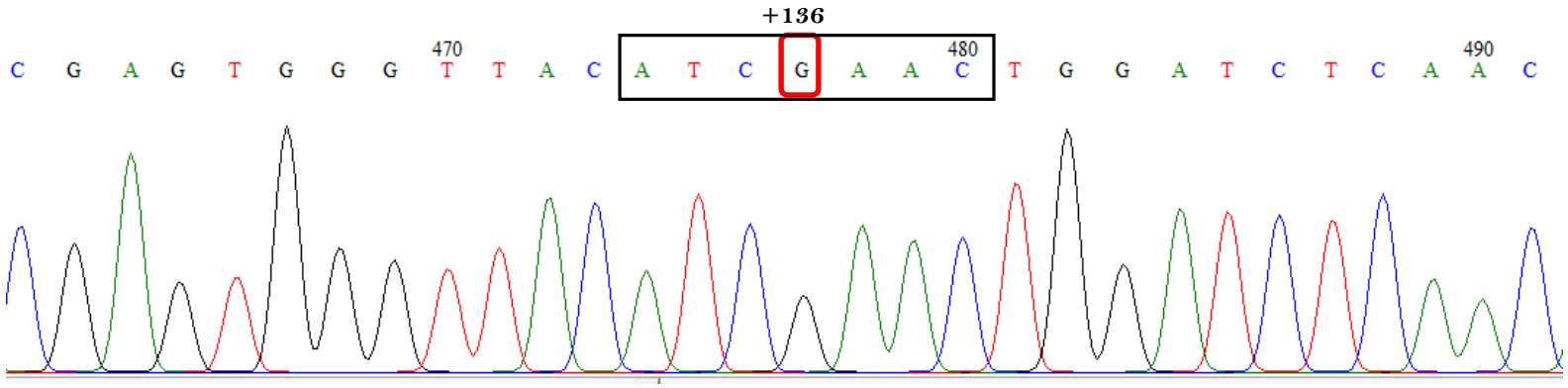
86. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.

87. Paquet, D., et al., *Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.* Nature, 2016. **533**(7601): p. 125-9.

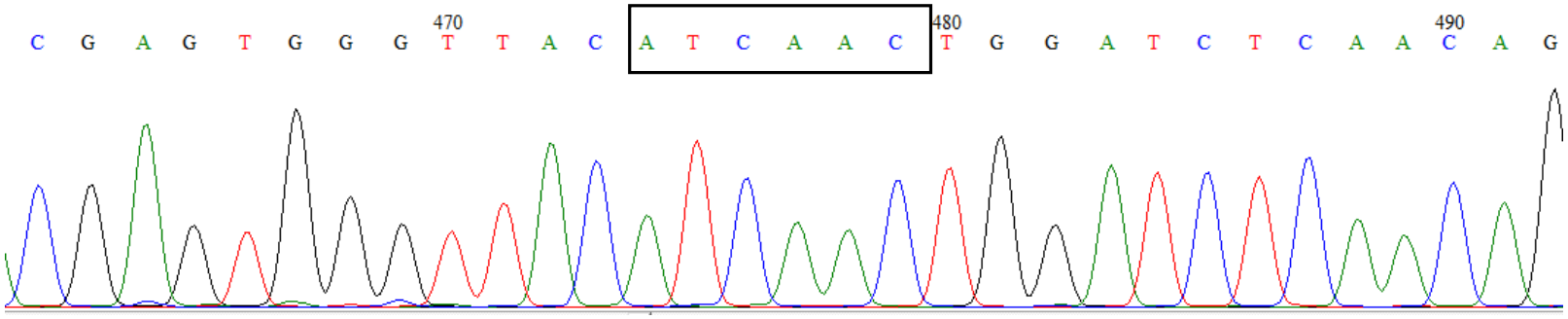
88. Zimmer, C.T., et al., *A CRISPR/Cas9 mediated point mutation in the alpha 6 subunit of the nicotinic acetylcholine receptor confers resistance to spinosad in Drosophila melanogaster.* Insect Biochem Mol Biol, 2016. **73**: p. 62-9.

Fig 1. The frameshift mutagenesis of the *bla* gene in the plasmid pBR322

1A



1B



1C

				136			
Wild-type (*bla* +):	**…A**	**T**	**C**	**G**	**A**	**A**	**C…**
	|	|	|		|	|	|
Frameshift (*bla* -):	**…A**	**T**	**C**	**-**	**A**	**A**	**C…**

Fig 2. Growth of different *E.coli* strains on ACPs and TCPs

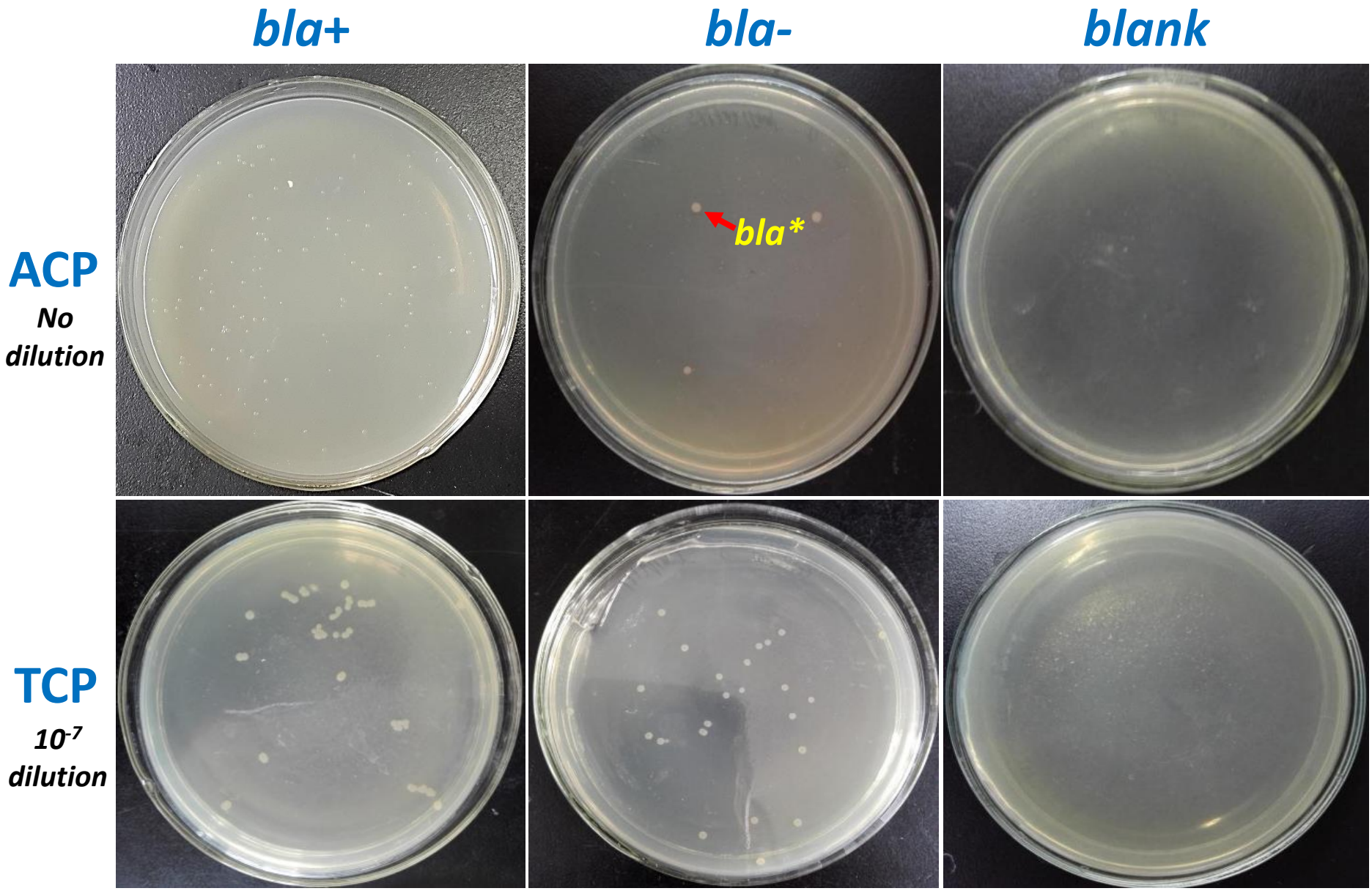


Fig 3A *Random mutagenesis*

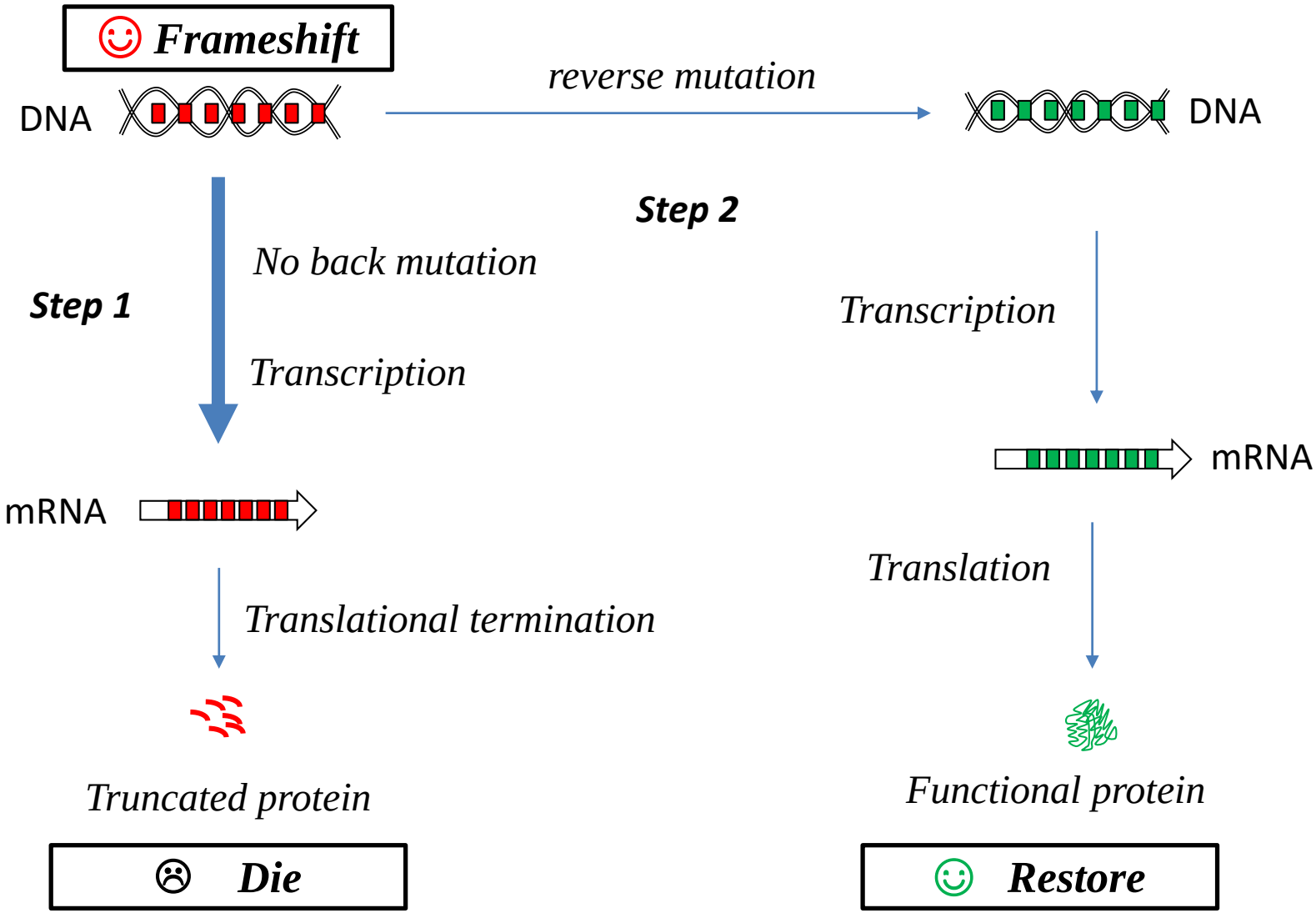


Fig 3B *Nonsense-mediated gene editing (NMGE)*

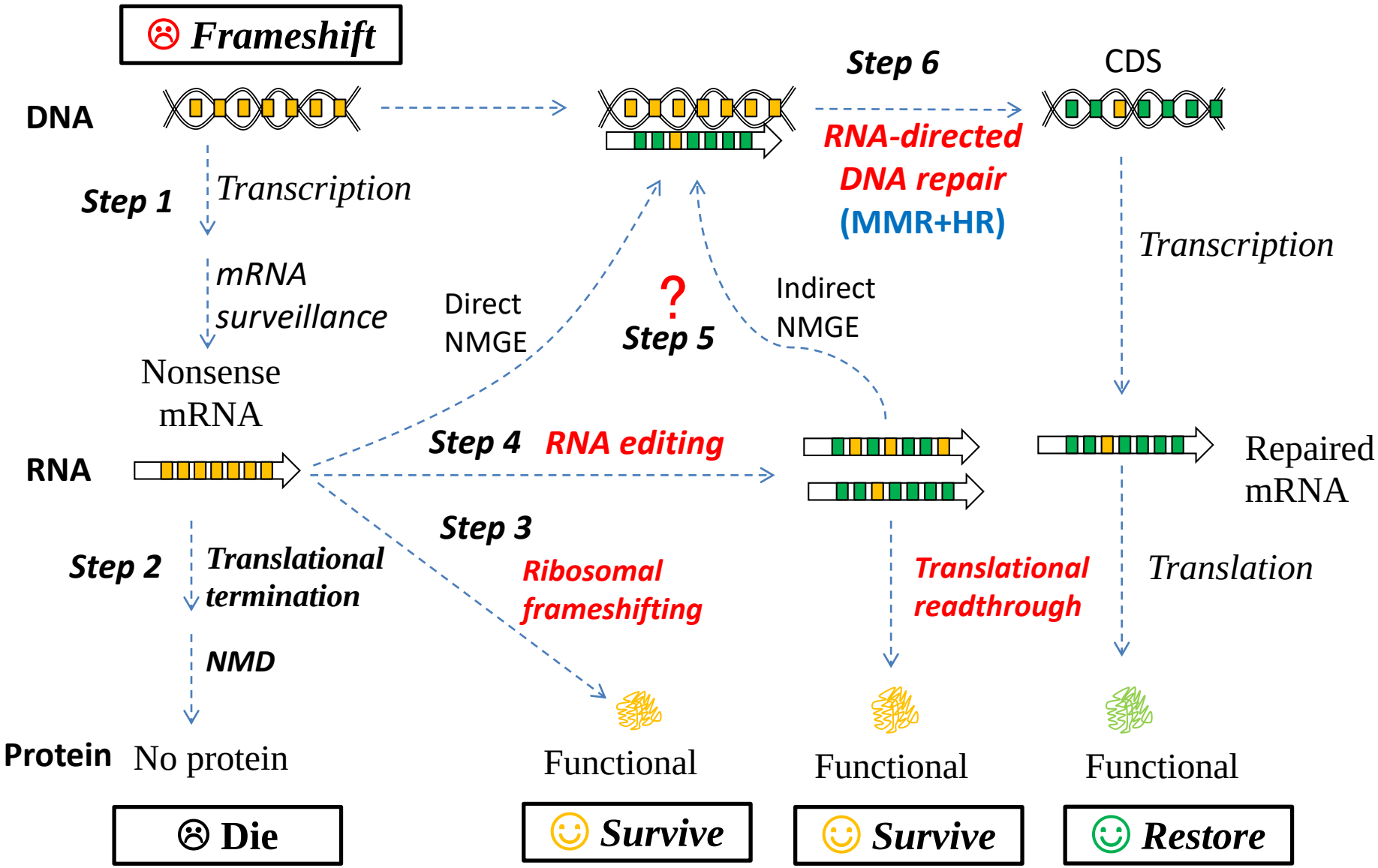


Fig 4A. The Sanger diagram of *bla**

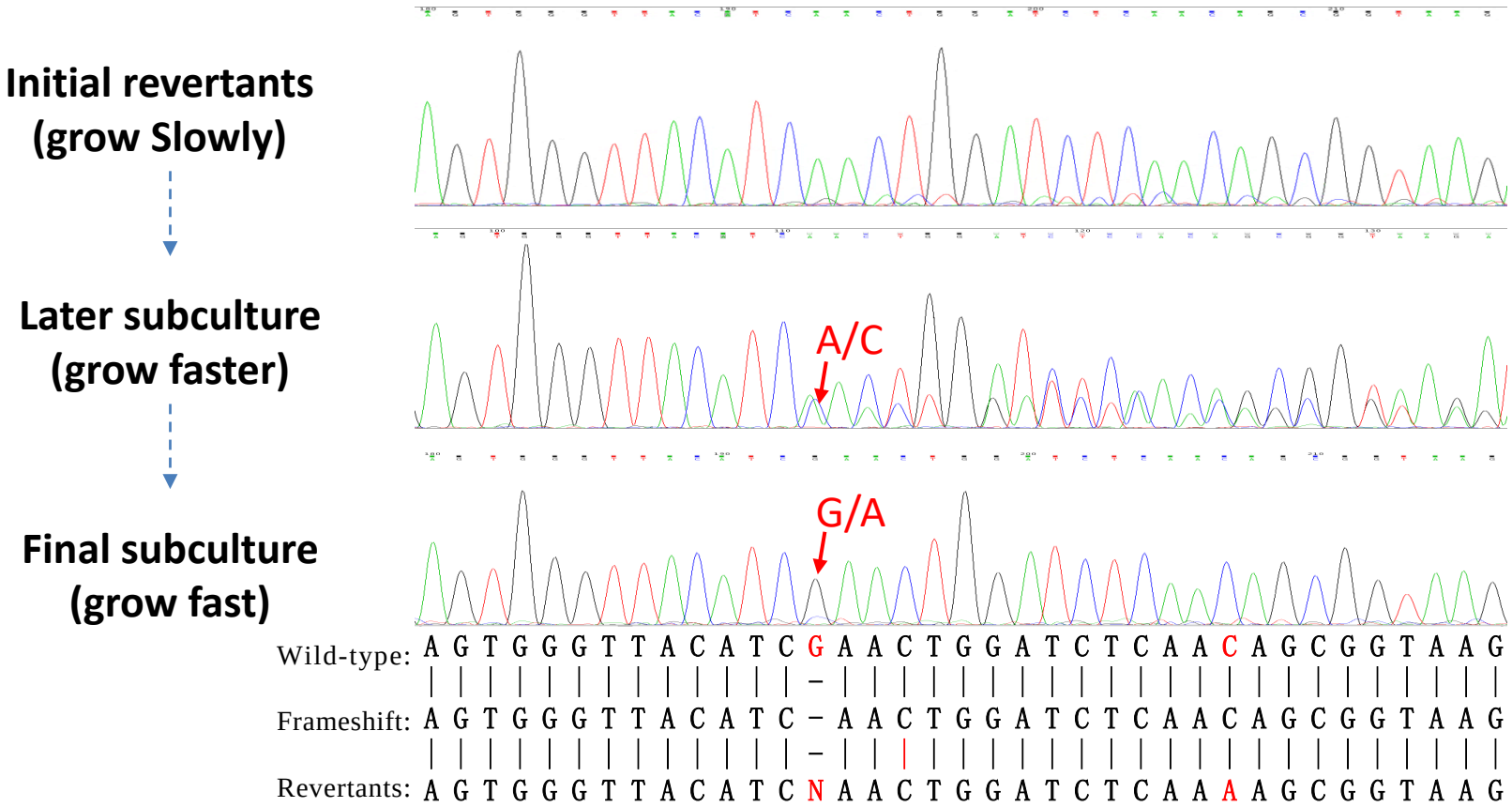


Fig 4B. The *bla* sequences of the wild-type and revertants

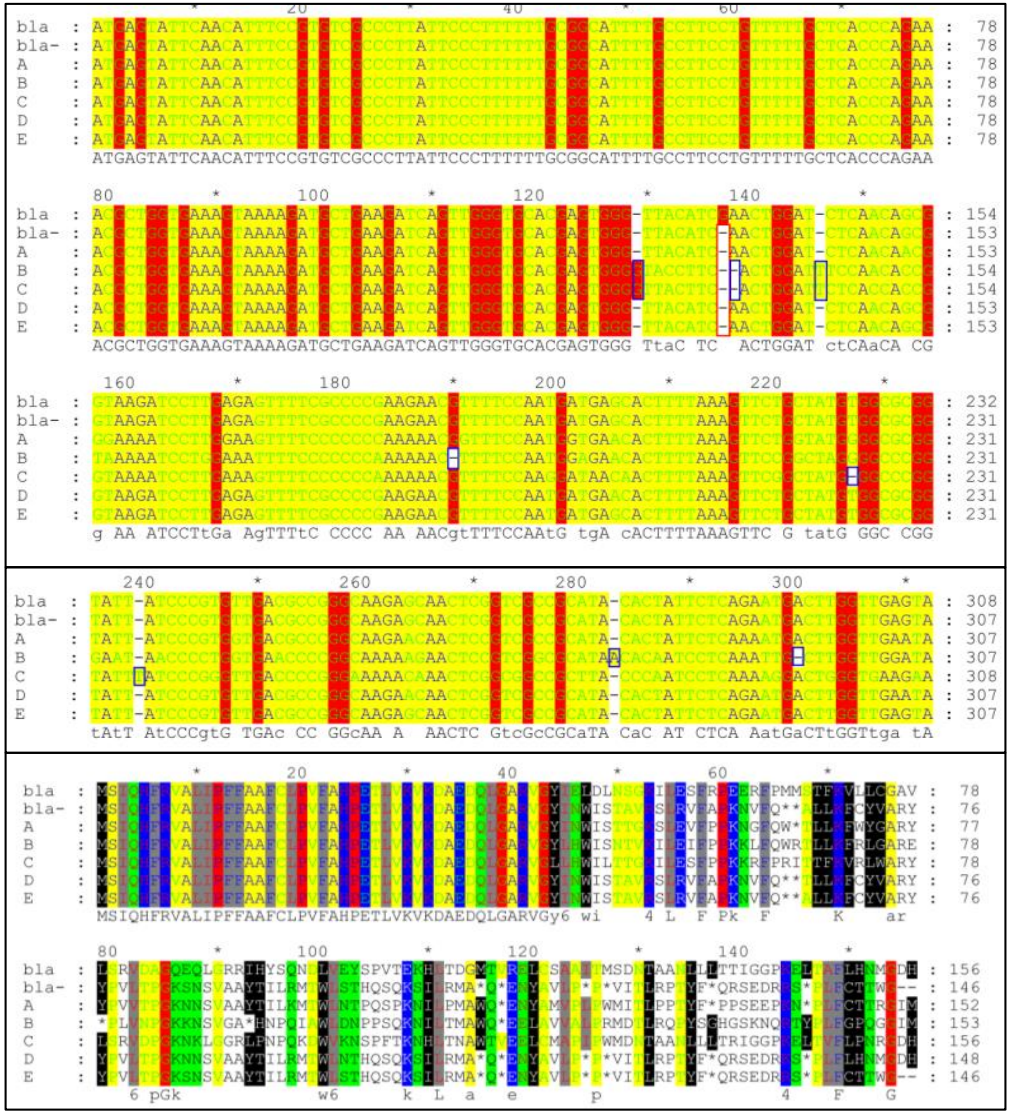


Fig 4C. The SNP/InDel in the genome of the frameshift and the revertant

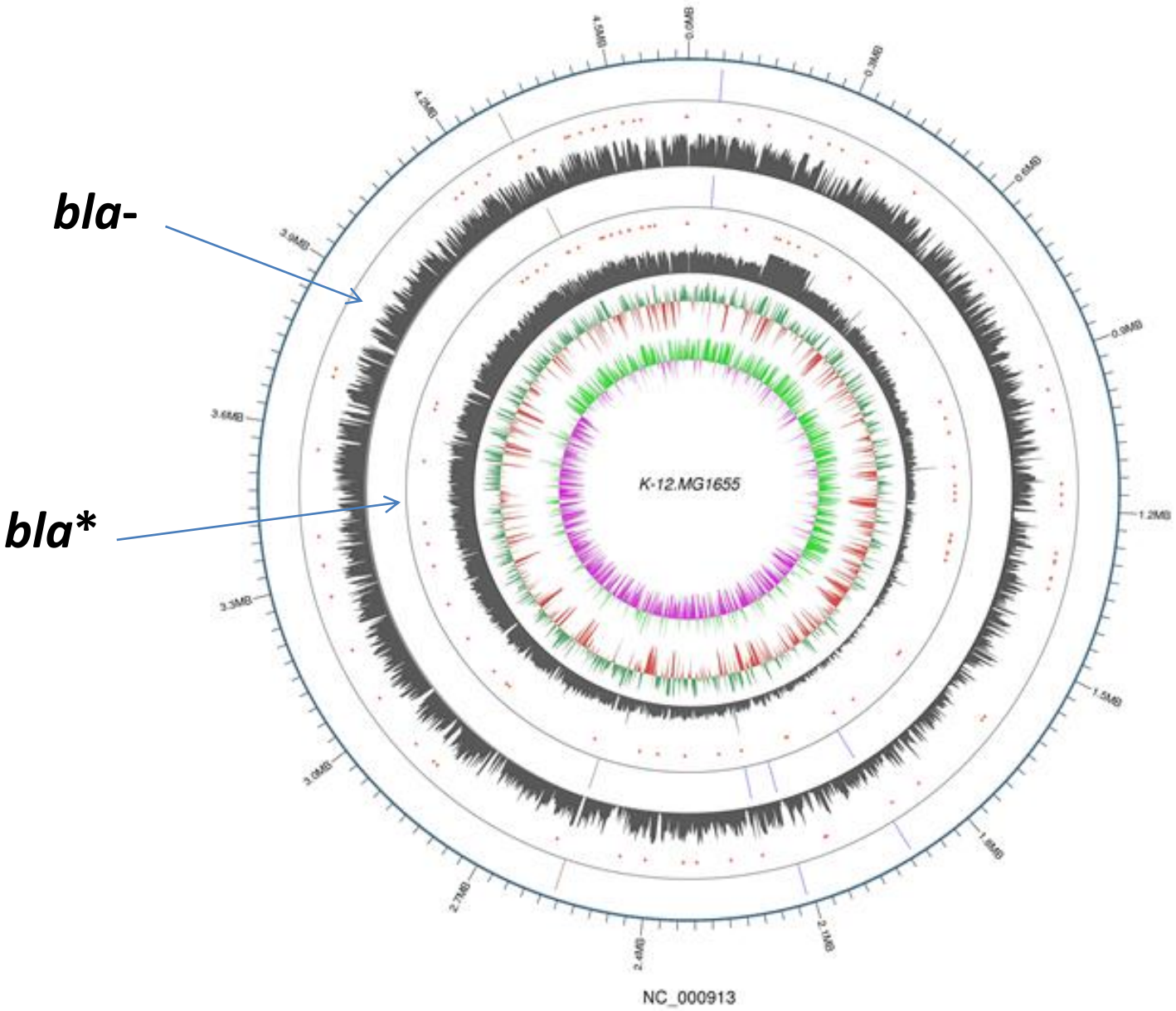


Fig 4D. The SNP/InDel in the genome of the frameshift and the revertant

bla-

bla*

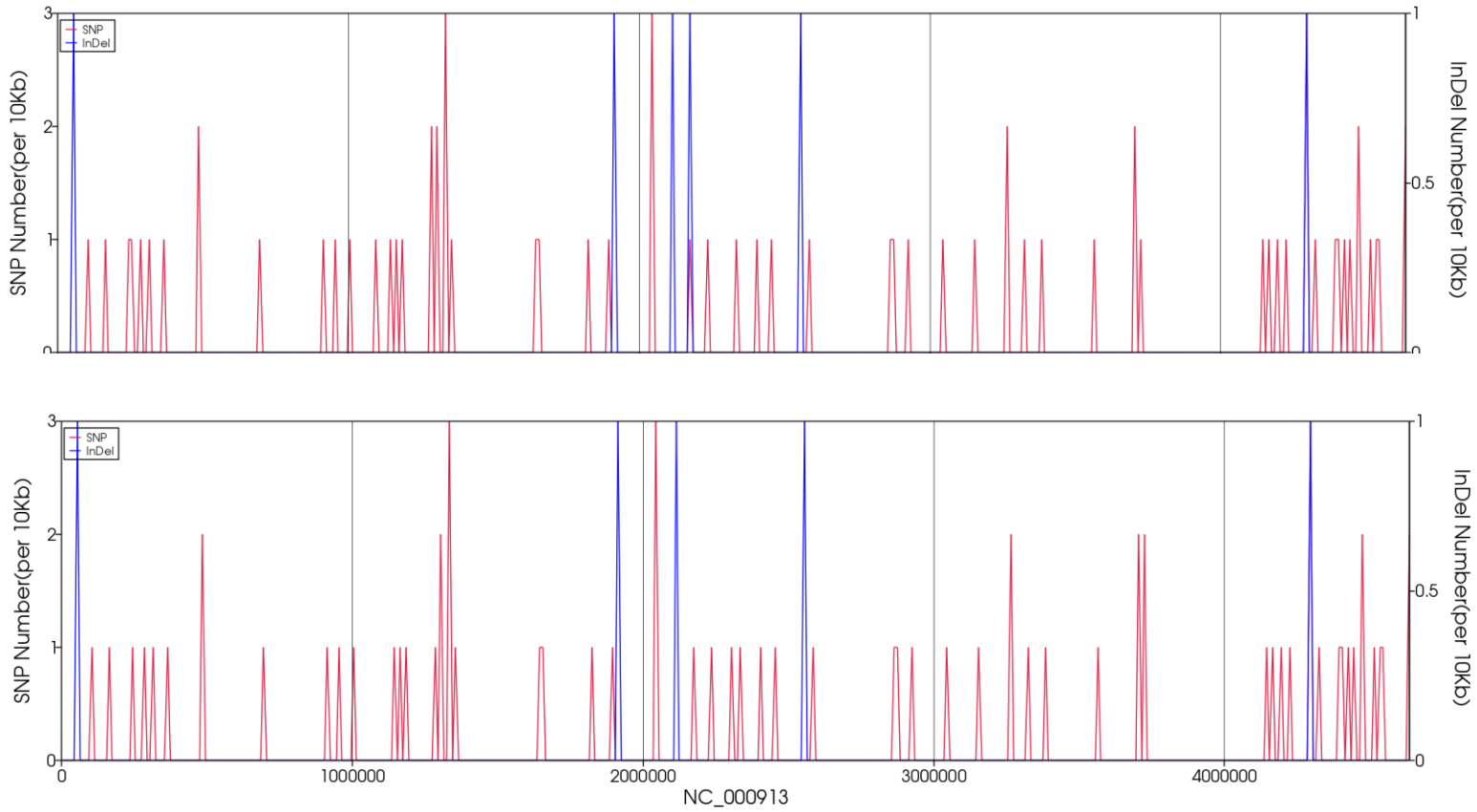


Fig 4E. The structure variation (SV) in the genome of the frameshift and the revertant

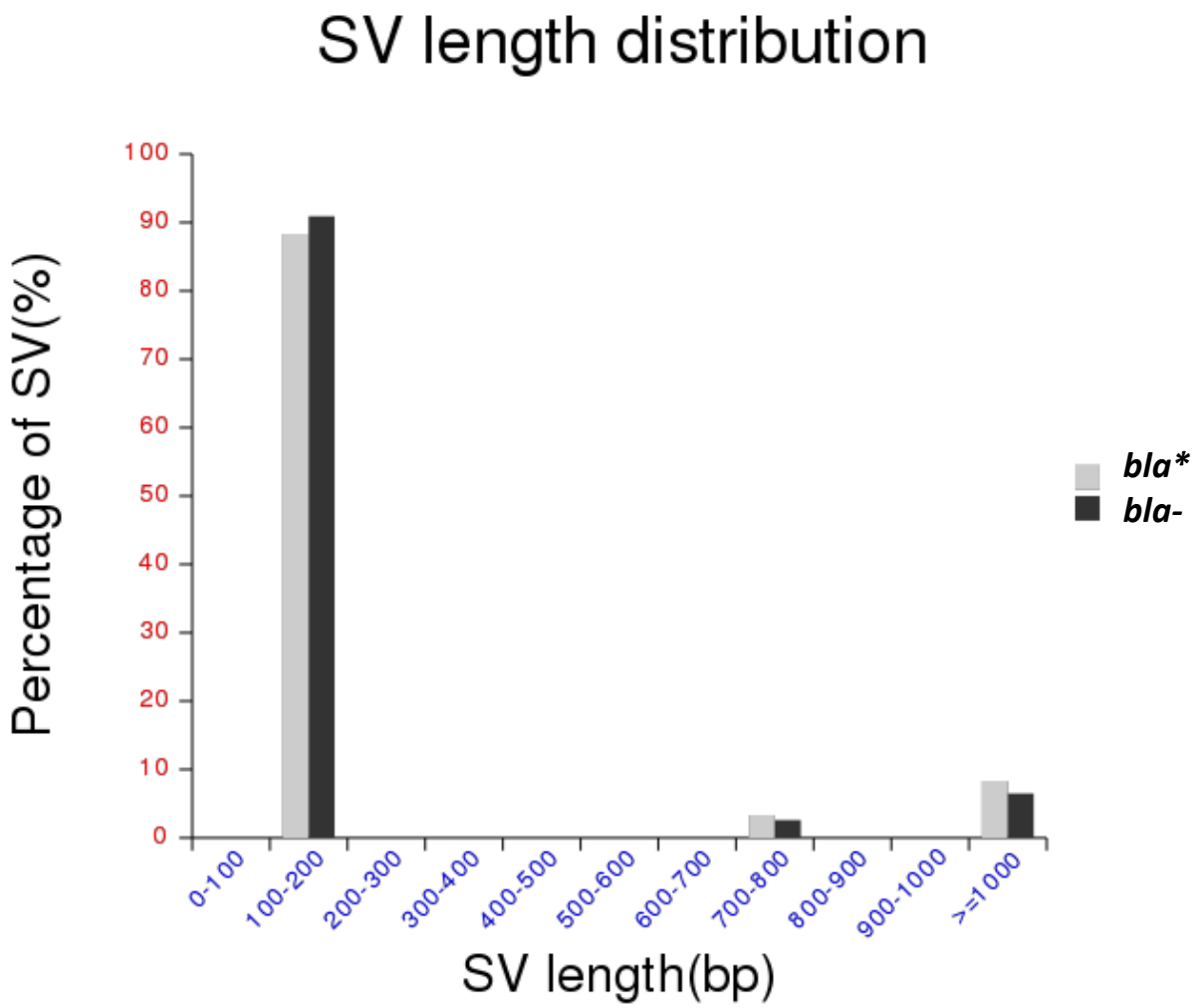


Fig 5. *Expression of the PTC-free frameshifted bla (bla#)*

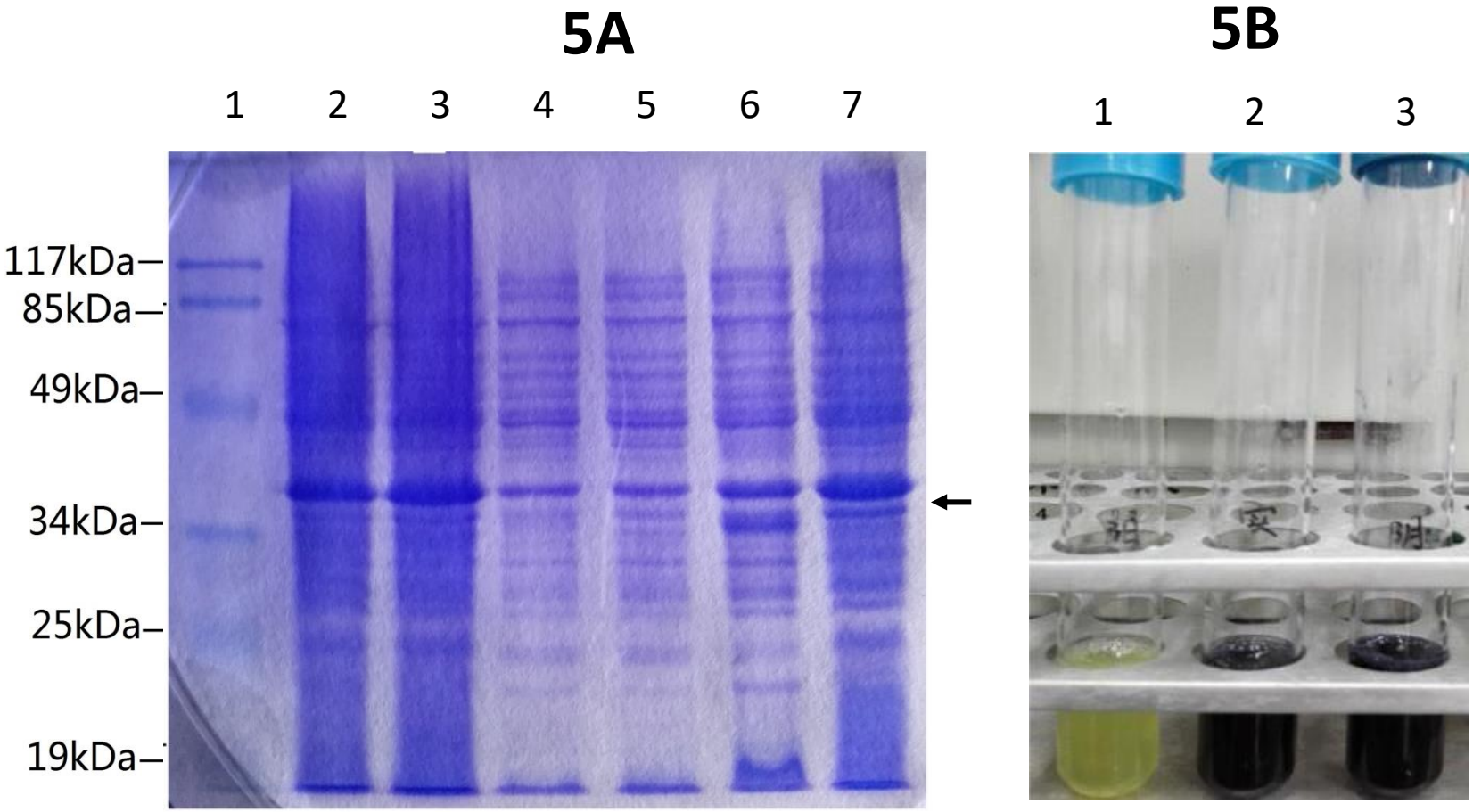
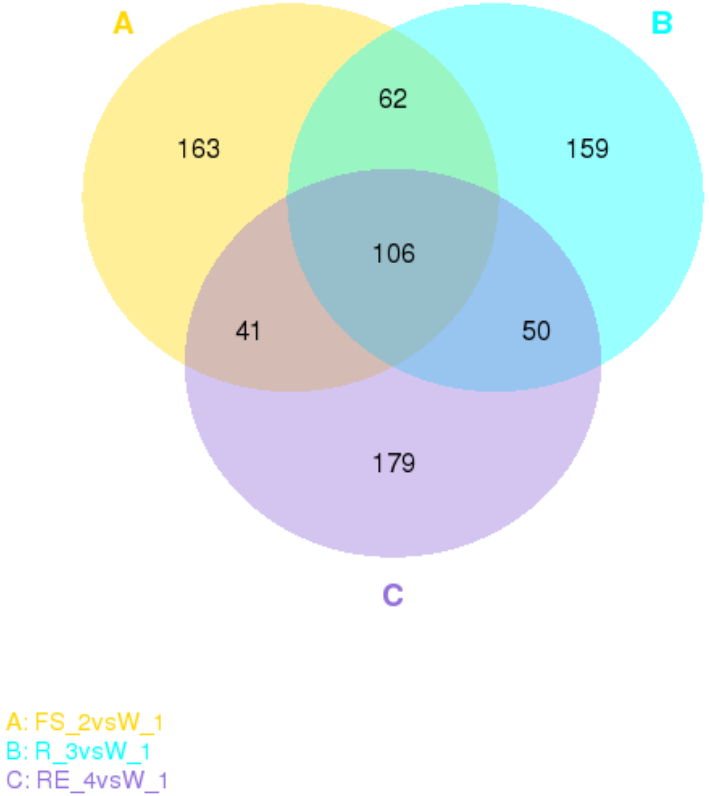
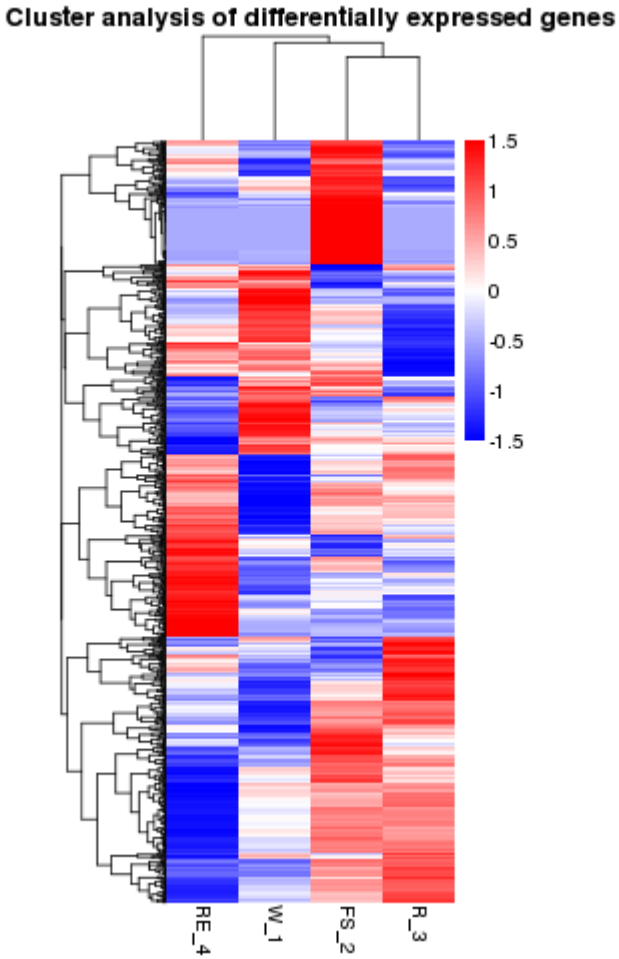


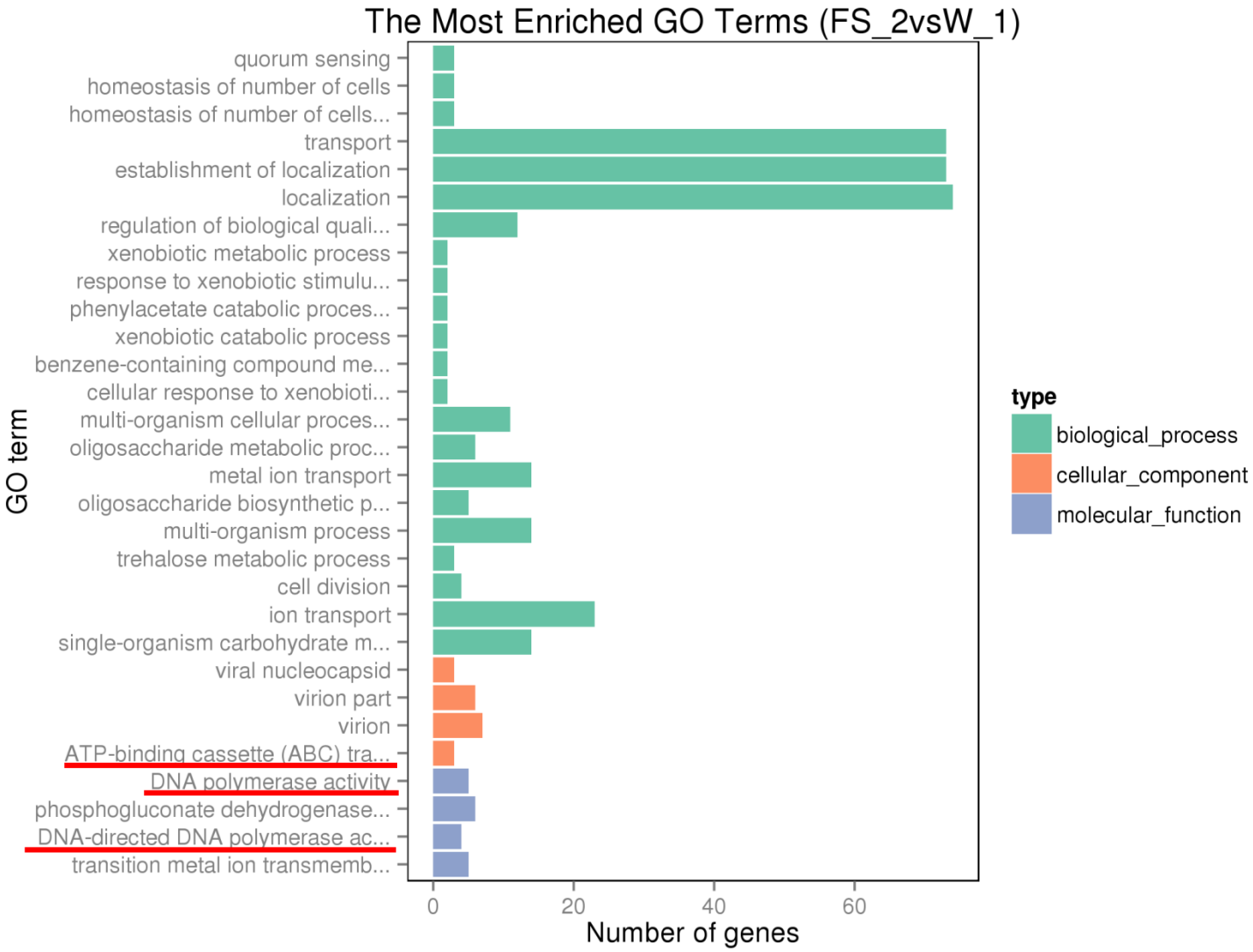
Fig 6. Transcriptome Analysis

6A Numbers of differential expression genes (DGE)

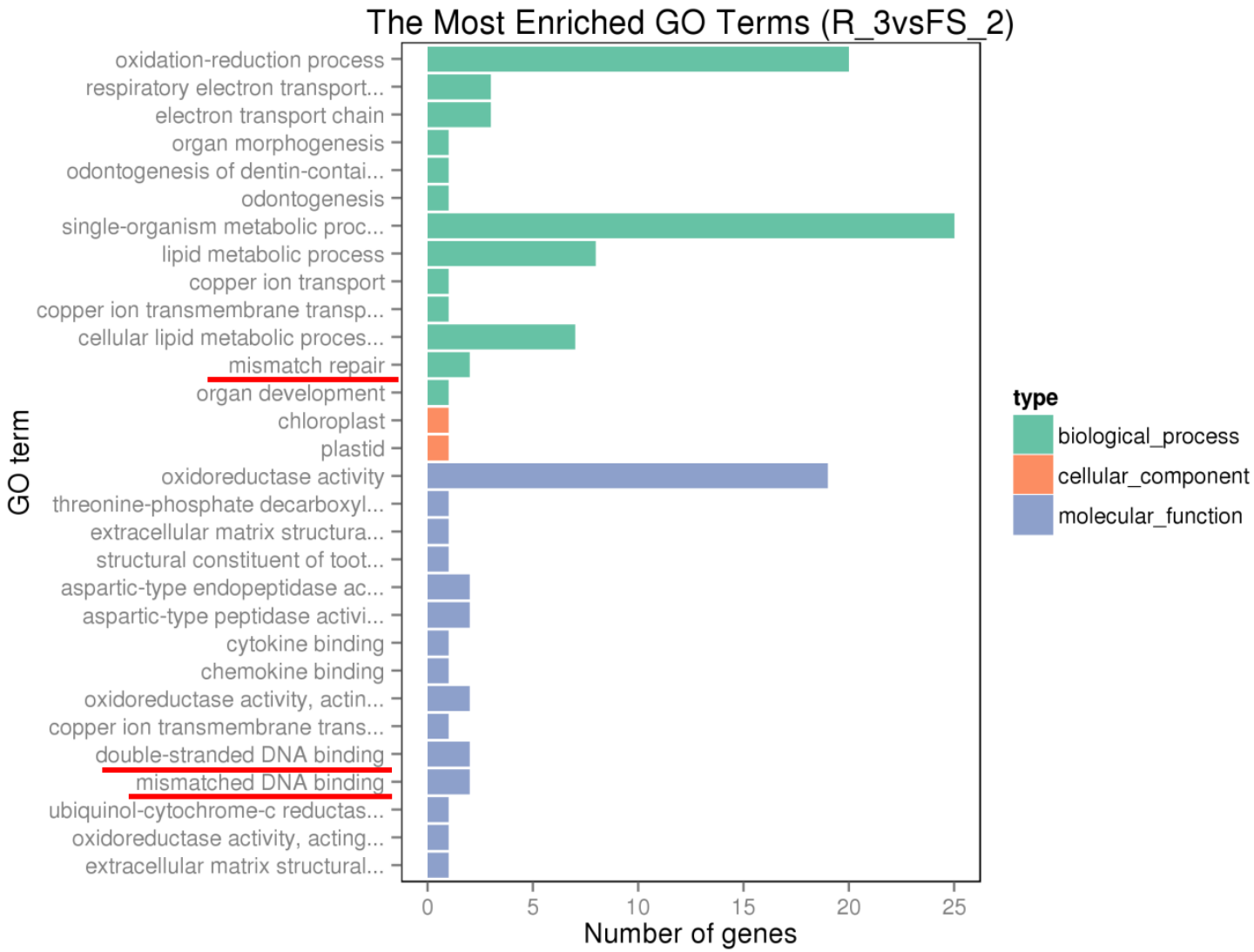




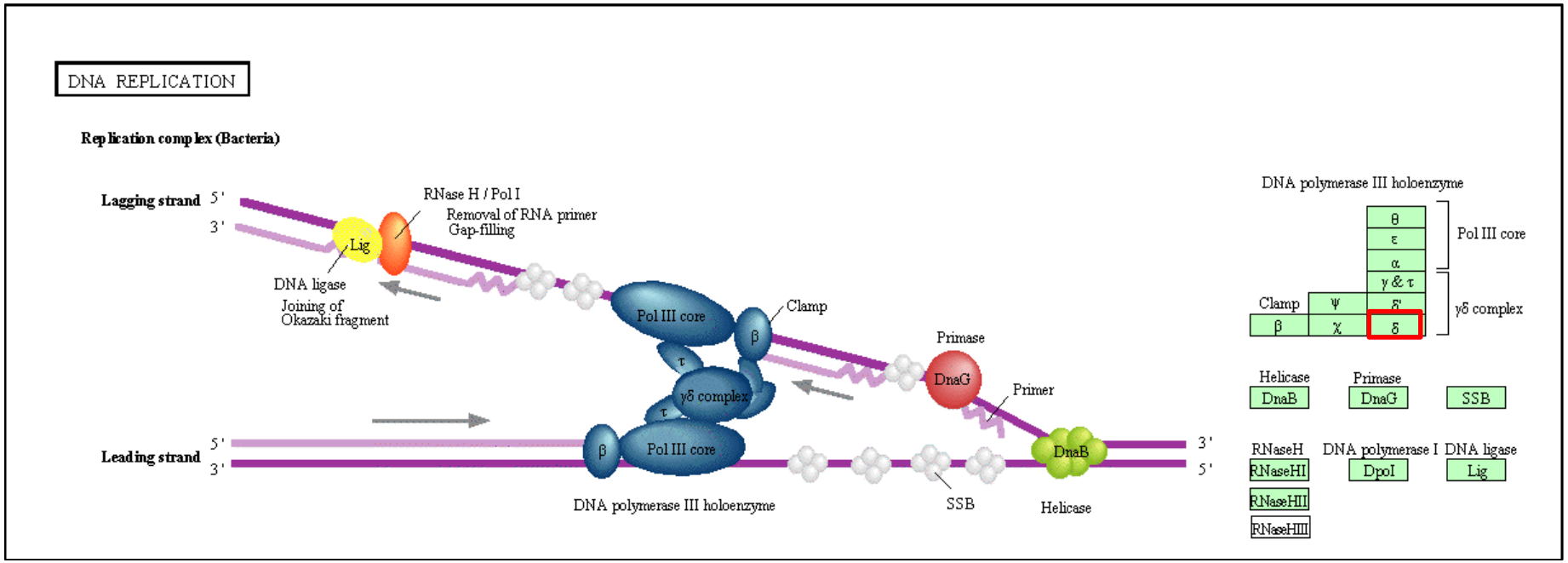
6B


The Most Enriched GO Terms (FS_2vsW_1)
GO term
quorum sensing
homeostasis of number of cells
homeostasis of number of cells...
transport
establishment of localization
localization
regulation of biological quali...
xenobiotic metabolic process
response to xenobiotic stimulu...
phenylacetate catabolic proces...
xenobiotic catabolic process
benzene-containing compound me...
cellular response to xenobioti...
multi-organism cellular proces...
oligosaccharide metabolic proc...
metal ion transport
oligosaccharide biosynthetic p...
multi-organism process
trehalose metabolic process
cell division
ion transport
single-organism carbohydrate m...
viral nucleocapsid
virion part
virion
ATP-binding cassette (ABC) tra...
DNA polymerase activity
phosphogluconate dehydrogenase...
DNA-directed DNA polymerase ac...
transition metal ion transmemb...
0
20
40
60
Number of genes
type
biological_process
cellular_component
molecular_function

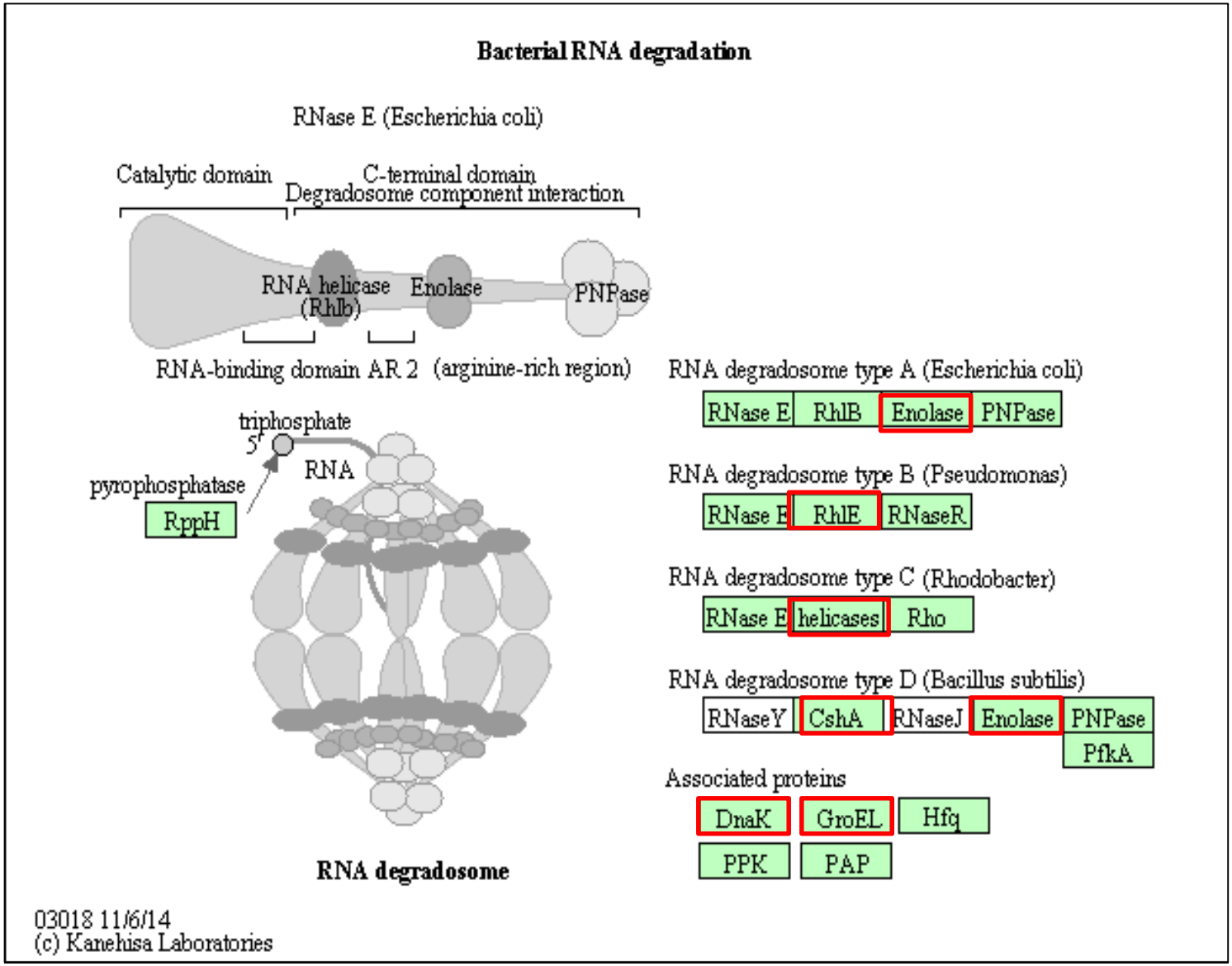
6C


The Most Enriched GO Terms (R_3vsFS_2)
oxidation-reduction process
respiratory electron transport...
electron transport chain
organ morphogenesis
odontogenesis of dentin-contai...
odontogenesis
single-organism metabolic proc...
lipid metabolic process
copper ion transport
copper ion transmembrane transp...
cellular lipid metabolic proces...
mismatch repair
organ development
chloroplast
plastid
oxidoreductase activity
threonine-phosphate decarboxyl...
extracellular matrix structura...
structural constituent of toot...
aspartic-type endopeptidase ac...
aspartic-type peptidase activi...
cytokine binding
chemokine binding
oxidoreductase activity, actin...
copper ion transmembrane trans...
double-stranded DNA binding
mismatched DNA binding
ubiquinol-cytochrome-c reductas...
oxidoreductase activity, acting...
extracellular matrix structural...
GO term
0
5
10
15
20
25
Number of genes
type
biological_process
cellular_component
molecular_function

6D genes upregulated in DNA replication



6E genes upregulated in RNA degradation/processing



6F genes upregulated in mismatch repair/homologous recombination

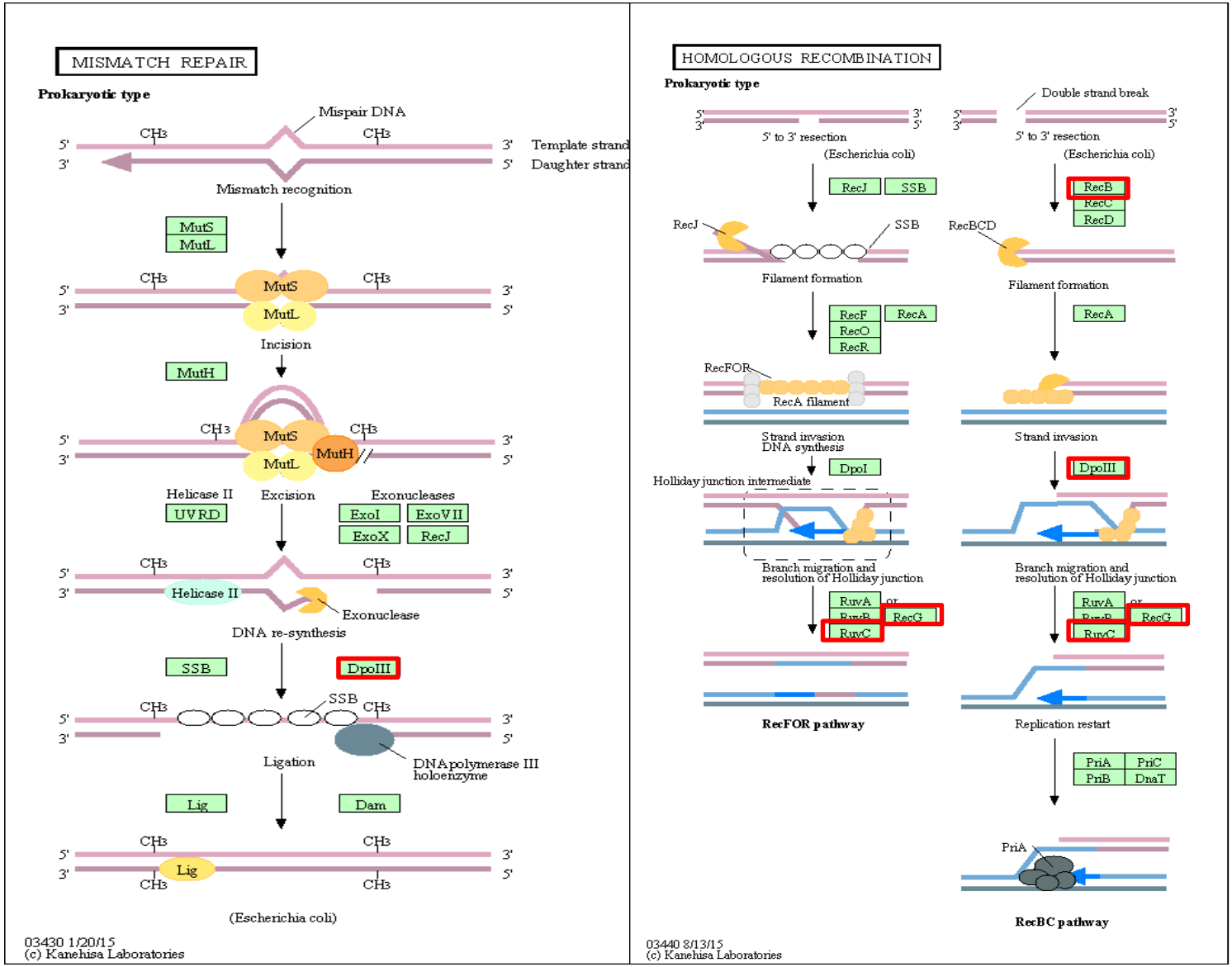


Fig 7A. Frameshift repair induced by DNA or RNA oligos (10 nM)

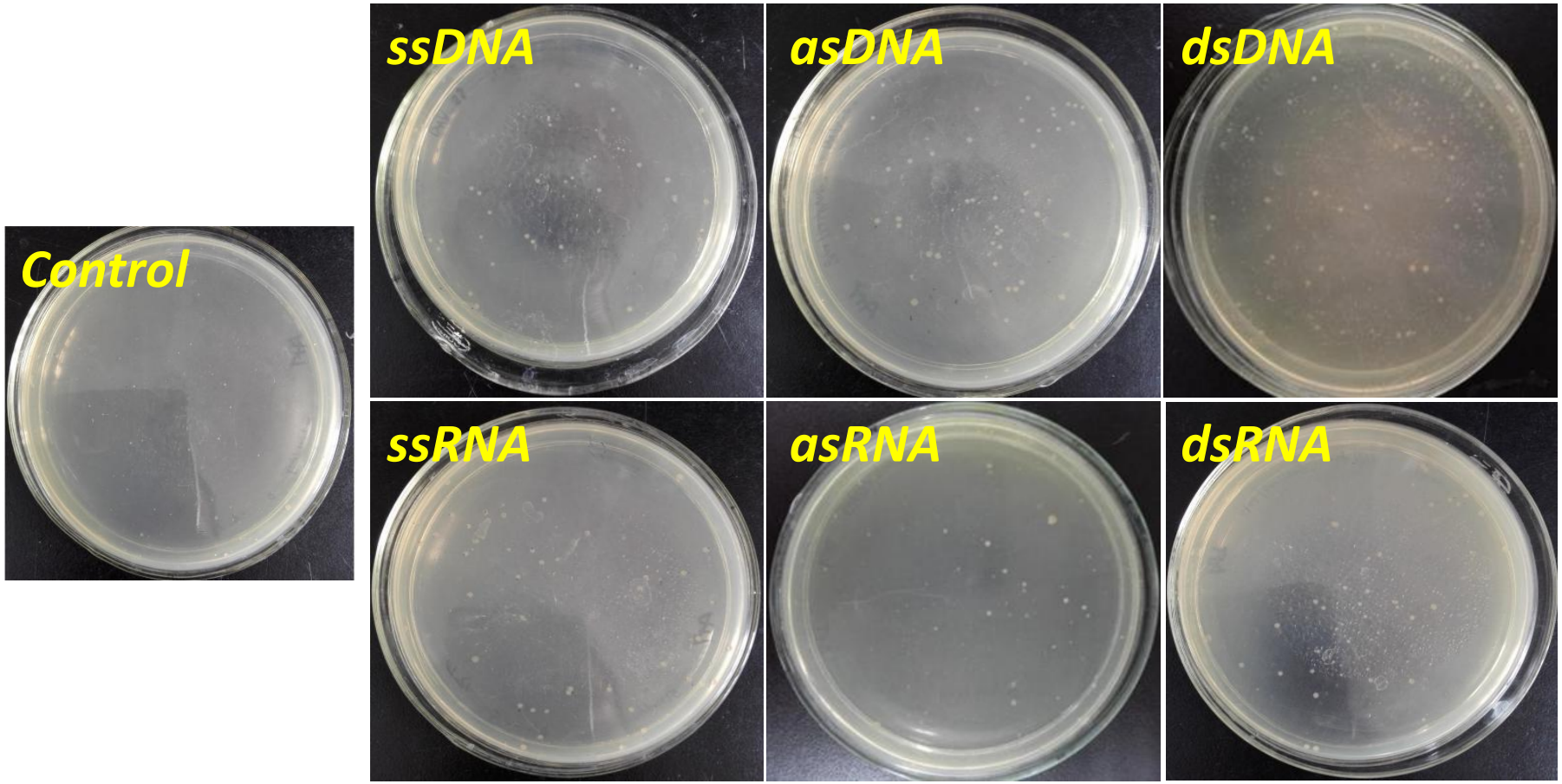


Fig 7B. The rate of reversions in different group

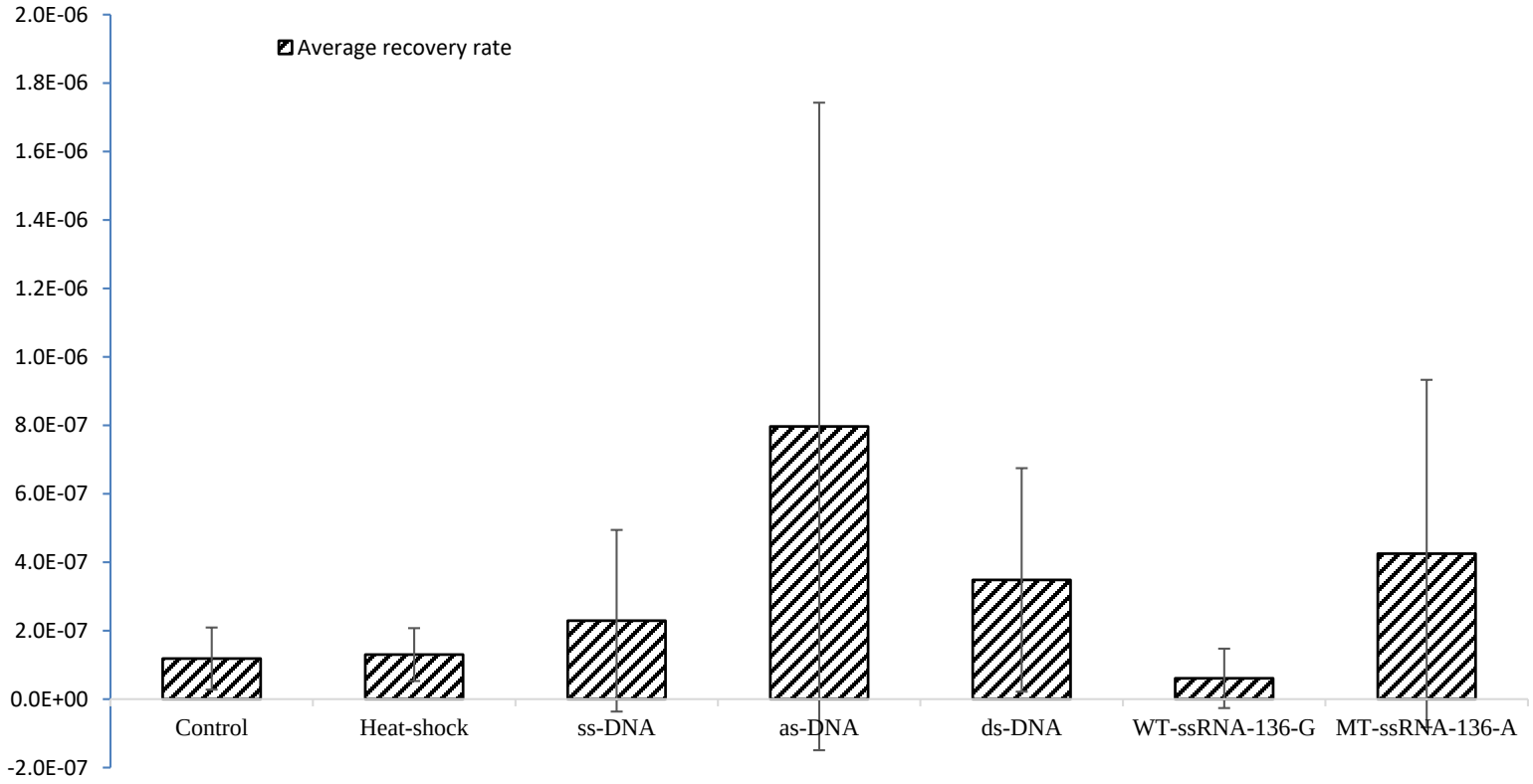
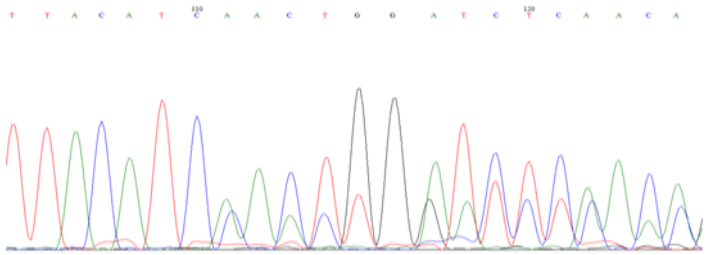


Fig 7C. Sanger sequencing diagram of the ssDNA-induced revertants



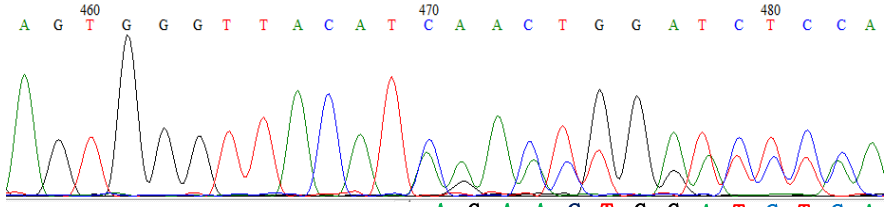
A A A C T G G A T C T C A A C A G C G G T

Main Peaks: T G C A C G A G T G G G T T A C A T C – A A C T G G A T C T C A A C A G C G G T

2nd Peaks: T G G A C G A G T G G G T T A C A T C A A A C T G G A T C T C A A C A G C G G T

WT DNA oligo: T G G A C G A G T G G G T T A C A T C G A A C T G G A T C T C A A C A G C G G T

Wild-type: T G C A C G A G T G G G T T A C A T C G A A C T G G A T C T C A A C A G C G G T



A G A A C T G G A T C T C A A C A G C G G T

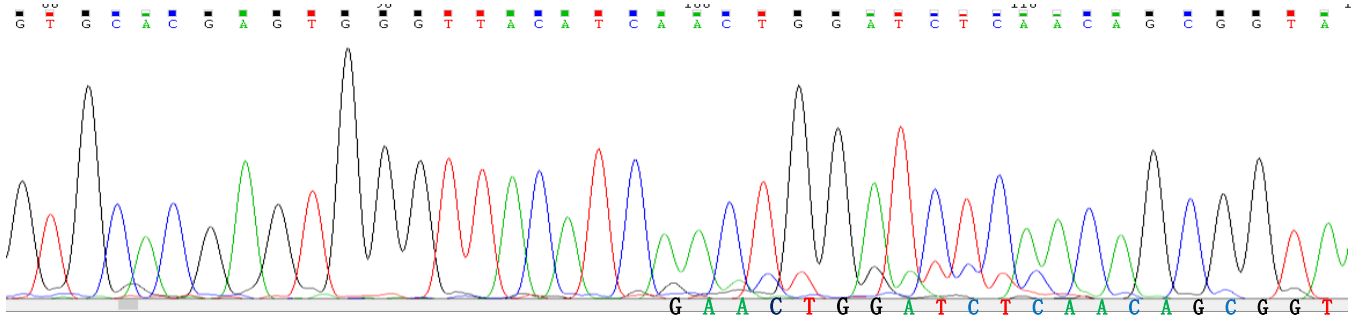
Main Peaks: T G C A C G A G T G G G T T A C A T C – A A C T G G A T C T C A A C A G C G G T

2nd Peaks: T G G A C G A G T G G G T T A C A T A G A A C T G G A T C T C A A C A G C G G T

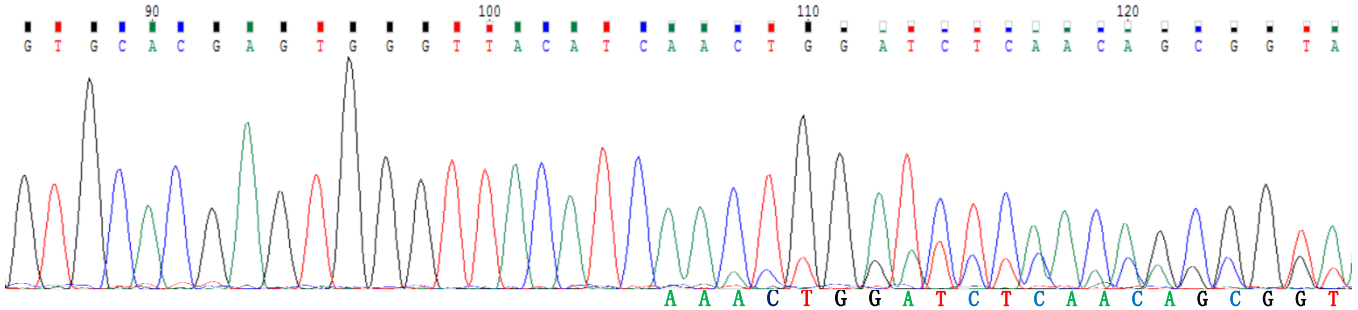
MT DNA oligo: T G G A C G A G T G G G T T A C A T A G A A C T G G A T C T C A A C A G C G G T

Wild-type: T G C A C G A G T G G G T T A C A T C G A A C T G G A T C T C A A C A G C G G T

Fig 7D. Sanger sequencing diagram of the RNA-induced revertants




```
Main Peaks:   TGCACGAGTGGGTTACATC-AACTGGATCTCAACAGCGGT
              |||||||||||||||||||  ||||||||||||||||||||
2nd Peaks:    TGGACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGT
              ||||||||||||||||||||||||||||||||||||||||
WT RNA oligo: UGGACGAGUGGGUUACAUCGAACUGGAUCUCAACAGCGGU
              ||||||||||||||||||||||||||||||||||||||||
Wild-type:    TGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGT
```




```
Main Peaks:   TGCACGAGTGGGTTACATC-AACTGGATCTCAACAGCGGT
              |||||||||||||||||||  ||||||||||||||||||||
2nd Peaks:    TGGACGAGTGGGTTACATCAAACTGGATCTCAACAGCGGT
              ||||||||||||||||||||||||||||||||||||||||
MT RNA oligo: UGGACGAGUGGGUUACAUCAAACUGGAUCUCAACAGCGGU
              ||||||||||||||||||||||||||||||||||||||||
Wild-type:    TGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGT
```