RESEARCH ARTICLE

Identification of potential TALEN and CRISPR/Cas9 targets of selected genes of some human pathogens which cause persistent infections

WGBP Dharmasena¹ and DHH Munasinghe^{1,2*}

¹Department of Botany, Faculty of Applied Sciences, University of Sri Jayewardenepura, Gangodawila, Nugegoda.

²Centre for Plant Materials and Herbal Products Research, University of Sri Jayewardenepura, Gangodawila, Nugegoda.

Submitted: Submitted: 09 July 2020; Revised: 15 April 2021; Accepted: 25 June 2021

Abstract: The human pathogens, Epstein Barr virus, human papilloma virus, herpes simplex virus-2, hepatitis B virus and *Leishmania* species can cause persistent infections, which cannot be cured with currently available treatments. The modern gene editing techniques, transcription activator like effector nuclease (TALEN) and clustered regularly interspaced short palindromic repeat / CRISPR associated protein 9 (CRISPR/Cas9), are potential candidates for their treatment. In this study, target sites for TALEN and CRISPR/Cas9 were identified *in silico* on selected essential and indispensable genes of the above pathogens, targeting the cease of the essential functions and curing the infection. The gene sequences of the pathogens were obtained from public databases and conserved sequences were identified. Then potential TALEN target sites were identified. For some selected targets, the off-target effects on the genomes of human, mouse, same pathogen and other organisms were tested and the putative functions of the mutated proteins were predicted. TALEN targets without having a potential off-target effect and not leading to mutated proteins with undesirable functions were selected for each gene. The potential CRISPR/Cas9 targets without off-target effect on human and murine genomes were identified and other off-target effects were evaluated. Results showed that potential TALEN and/or CRISPR/Cas9 targets with higher binding specificity and efficiency were available for the selected genes. It can be concluded that the selected targets can potentially be used to produce respective proteins, and *in vitro* and *in vivo* applications are potentially possible.

Keywords: Bioinformatics, CRISPR/Cas9, genome editing, persistent pathogens, TALEN

INTRODUCTION

Genome editing is a technique that is used to alter the genomes of organisms. There are mainly three types of genome editing techniques known as zinc finger nuclease (ZFN), transcription activator like effector nuclease (TALEN) and clustered regularly interspaced short palindromic repeat / CRISPR associated protein 9 (CRISPR/Cas9). All these are site - specific nucleases (Ochiai Yamamoto, 2015). Among them ZFN and TALEN are protein based chimeric nucleases while CRISPR/Cas is RNA based (Nemudryi *et al.*, 2014). TALEN and CRISPR/Cas9 are of greater interest due to high specificity of binding (Yeadon, 2014). These techniques have recently been used in the gene editing of organisms such as zebra fish (Gonzales & Yeh, 2014), *Arabidopsis thaliana* (Feng *et al.*, 2013), *Zea mays* (Char *et al.*, 2015; Kelliher *et al.*, 2017), *Oryza sativa* (Shen *et al.*, 2017; Han *et al.*, 2019; Usman *et al.*, 2021) and of human cell lines (Yuen *et al.*, 2015).

TALEN and CRISPR/Cas9 techniques have a high potential for application as alternative treatment strategies. The CRISPR/Cas9 method has been tested to cure some diseases such as β thalassemia (Xu *et al.*, 2015; Frangoul *et al.*, 2021), cystic fibrosis (Schwank *et al.*, 2013; Fan *et al.*, 2018), hemophilia A (Park *et al.*, 2015), hepatitis B (Seeger Sohn, 2014) and human

* Corresponding author (helani@sci.sjp.ac.lk <https://orcid.org/0000-0001-7229-4691>)

This article is published under the Creative Commons CC-BY-ND License (<http://creativecommons.org/licenses/by-nd/4.0/>). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

immunodeficiency virus (HIV) infections (Liu *et al.*, 2017). TALEN technology has been tested to cure Duchenne muscular dystrophy (Li *et al.*, 2015), HIV (Hutter *et al.*, 2015; Benjamin *et al.*, 2016) and human papilloma viral infections (HPV) (Shankar *et al.*, 2017).

Since these techniques are based on site - specific nucleases, specific sites to be bound to the nucleases can be previously determined. The essential or indispensable genes for infection, pathogenesis or persistence of the pathogens can be targeted and these genes could be mutated to cure the infection or to diminish the pathogens' activities. The techniques are highly effective to be used in persistent incurable infections or that have developed resistance against currently available drugs. Some examples of causative agents of these persistent infections are herpes simplex virus (HSV), hepatitis B virus (HBV), HPV, HIV, Epstein Barr virus (EBV), cytomegalovirus (CMV), some bacteria, fungi and protists (Boldogh *et al.*, 1996).

The site where the TALEN or CRISPR/Cas9 binds in a specific gene can be predicted, according to that the TALEN or CRISPR/Cas9 proteins can be designed and synthesised. The main criterion in designing a TALEN is the positioning of a Thymine (T) nucleotide at the 5' end of the target. In CRISPR/Cas9, the presence of Protospacer Adjacent Motif (PAM) is the essential criterion (Nemudryi *et al.*, 2014).

The objective of this study was to identify TALEN and CRISPR/Cas9 potential targets and to predict their off-target effects to select the best target sites for some selected genes of EBV, HBV, HPV, HSV-2 and *Leishmania donovani* and *Leishmania infantum* pathogens, which are showing persistent infection and for which treatment is not currently available. The genes of these pathogens were selected with the aim to cease the pathogen persistency and replication. In EBV, the gene *LMP2A* is essential for the persistence of the virus (Longnecker, 2000), and *EBNA1* gene is essential for the replication and transcription (Sivachandran *et al.*, 2012). It can be postulated that the introduction of a site-specific nuclease for *LMP2A* gene followed by that of *EBNA1* would diminish the viral content in the host leading to eradication of the virus with continual application. In the same way, the *UL21* and *UL30* of HSV-2 could be mutated which are essential in viral propagation (Le Sage *et al.*, 2013) and replication (Liu *et al.*, 2006), respectively. The replication of the HPV can be inhibited by mutating the *E2* gene (Sanders & Stenlund, 2000; McBride, 2013) and also the oncogenic effect (Leykauf *et al.*, 2008) could be diminished. *HBx* gene of HBV is indispensable for the development of the viremia and persistency (Tsuge *et al.*, 2010) and its mutations might cease the viral infection in the host. *Leishmania*

sp., a pathogen that causes persistent infection could be controlled by mutating the *tryR* gene that protects against oxidative stress (Paul *et al.*, 2014).

METHODOLOGY

Selection of pathogen specific genes

The genes *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *UL21* and *UL30* of HSV-2, *HBx* of HBV and *tryR* of *L. donovani* and *L. infantum* were selected and subjected for TALEN and/or CRISPR/Cas9 target identification. Randomly selected entries were obtained from the databases 'GenBank' (Clark *et al.*, 2016) and 'RefSeq' (O'Leary *et al.*, 2015). These sequences were tested using the tool 'NCBI conserved domains' (Marchler-Bauer *et al.*, 2014) and from the results, the sequences that confirm the presence of the gene were selected. Then the open reading frame (ORF) responsible for the gene in each sequence was identified using the tool 'ORF Finder' (Rombel *et al.*, 2002). The maximum length ORF of each sequence was obtained for the analysis.

Identification of conserved residues of the gene

Conserved residues of each selected gene were identified using the software 'Unipro UGENE' (Okonechnikov *et al.*, 2012) by aligning the maximum length ORFs of selected sequences. Then some conserved residues above 60 nucleotides were selected for identification of potential TALEN and CRISPR/Cas9 targets.

Identification of TALEN target sites

The tool 'TALEN Targeter' was used to identify the TALEN target sites. The selected conserved sequences of a selected gene were used as the input data. The predesigned TALEN architecture by Miller *et al.* (2011) was used to design TALEN targets and 'NH' was selected as the G substitute repeat variable diresidue (RVD). Then the parameters were adjusted to hide redundant TALENs in output. Other than these, guidelines by Streubel *et al.* (2012) were applied in the analysis. From the output, several TALEN targets having highest percentage of 'HD or NH' RVDs in the respective TALENs and having at least one unique restriction site at the spacer region were selected for further analyses. This procedure was followed for all selected conserved sequences in selected genes of *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *HBx* of HBV and *tryR* of *L. donovani* and *L. infantum*.

Identification of the potential off-target effect of the respective TALENs of the selected TALEN targets

The potential off-target effect or target specificity of the selected TALENs were identified for human genome and murine genome using two bioinformatic tools, TAL Effector Nucleotide Target 2.0 (Doyle *et al.*, 2012) and PROGNOS (Fine *et al.*, 2013). The RVD sequences of the TALEN targets selected above were used as the input data for both tools.

Apart from these the probable unnecessary bindings on the genome of the selected pathogen was also determined using the tool Paired Target Finder. There, a genome sequence of the pathogen in 'RefSeq' database was used as the target sequence and the RVD sequences of the selected TALENs as the query sequence.

Identification of potential off-target effect of TALENs on other organisms

Basic local alignment search tool (BLAST) (Altschup *et al.*, 1990) was used to identify the off-target effect of 'TALENs respective to the selected TALEN targets' in genomes of other organisms. Three methods were followed in the procedure. In the first method, TALEN target sequence was entered as the query sequence, where the spacer region was in lowercase letters, to the nucleotide BLAST tool. Filters were selected to mask lowercase letters and the search was carried out keeping all the other parameters default. In the second method instead of entering the whole TALEN target sequence as the query, only the TALE regions were entered as a continuous sequence by removing the spacer region. Next the same procedure as mentioned above was carried out except masking for lowercase letters. In the third method, all the nucleotides of the spacer region were replaced by the letter 'N' and used as the query sequence. The rest of this procedure is same as of the second method.

The suspected sequences for having off-target effect from the BLAST results were tested again using the tool 'Paired Target Finder' entering the NCBI accessions of the suspected sequences as the target and RVD sequence of the respective TALEN target as the query.

Identification of putative functions of the mutated protein

A sequence of a selected gene (Supplementary Table 1) was obtained and first, the nucleotides of the sequence

were numbered from 5 to 3 end using the Group DNA option of the tool Sequence Manipulation suite (Stothard, 2000).

Then the probable cut site by the first TALEN target was marked in the sequence and one nucleotide adjacent to the cut site was deleted. The resulting sequence was filtered to remove unnecessary numbering and spaces using the option Filter DNA of the tool Sequence Manipulation Suite. Then the ORFs in the sequence were identified using the tool 'ORF Finder'. After that, the ORFs responsible for amino acid sequences greater than 75 amino acids and which passed through the cut site were obtained. Each of these sequences was BLAST searched in the Protein BLAST tool (Altschup *et al.*, 1990). The same procedure was followed by deleting two nucleotides adjacent to the cut site. Then for all the other potential TALEN targets of the same gene and other selected genes, the same procedures were followed.

Identification of potential CRISPR/Cas9 target sites and their potential off-target effects

Potential CRISPR/Cas9 target sites were identified and off-target effect was predicted using the tool CCTop (Stemmer *et al.*, 2015). First, a conserved residue of a selected gene of the selected pathogen was entered into the tool as a plain text as the input. Then the maximum mismatches that an off-target should possess were set as four and the human genome (*Homo sapiens* GRCH38/hg38) was selected to identify off-targets. Other categories were kept default and submitted for analysis. In the same way, off-targets in the murine genome were identified by selecting the mouse genome (*Mus musculus* GRCm38/mm10). Then the above procedures were carried out for all the selected conserved residues of the genes *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *UL21* and *UL30* of HSV-2, *HBx* of HBV and *tryR* of *L. donovani* and *L. infantum*.

The tool CCTop displays the CRISPR/Cas9 target sequences of the query sequence in the order of off-targets in the selected genome from targets with null off-target effect to the targets with the highest off-target effect. Among them, the targets with the null off-target effect on both human and the murine genome were selected and were overlaid to identify the targets common to both the human and murine genomes with null off-target effect. The binding efficacies of the designed CRISPR/Cas9 nucleases were determined using the tool CRISPRator (Labuhn *et al.*, 2018).

Identification of the potential off-target effect of selected CRISPR/Cas9 targets on genomes of other organisms

The potential CRISPR/Cas9 target sites identified above were used as the query and BLASTN searches were carried out. From the results, the targets showing the off-target effect on other genomes were identified.

RESULTS AND DISCUSSION

Obtaining the ORF of confirmed gene sequence

The presence of the selected genes, *LMP2A* and *EBNA1* of EBV, *E2* of HPV type 16, *UL21* and *UL30* of HSV-2, *HBx* of HBV and *tryR* of *Leishmania* species in obtained sequences were confirmed from the results of the tool NCBI Conserved Domains. The ORF responsible for the gene in each sequence was identified from the results of the tool ORF Finder (Supplementary Table 2). The confirmation of a gene sequence is important because in some instances annotation errors are present in the sequences available in the NCBI database. The identification of ORF of the selected gene sequence is also of immense importance because a sequence obtained from the databases may contain areas that do not belong to the ORF of the gene. If such regions are present in the sequence, TALEN and/or CRISPR/Cas9 target sites may be identified for those regions too.

Selection of conserved residues to identify TALEN and CRISPR/Cas9 target sites

According to the selected TALEN architecture, the maximum length of a TALEN target site is 60 nucleotides. Therefore, from the gene sequence alignment results (Supplementary Table 3), the conserved sequences with minimum length of 60 nucleotides were selected for each gene (Supplementary Table 4). The same conserved sequences were used for the identification of CRISPR/Cas9 target sites. HPV types 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 70 were initially selected because these types are the high-risk types for cancer (Burd, 2003). But any conserved domain greater than 60 nucleotides were not observed in them. Then HPV type 16 was considered because it is the type that is responsible for the highest percentage of cancers among the HPV types (National Cancer Institute of USA, 2017). For *Leishmania* sp. also any conserved sequence in enough length for a TALEN target was

not identified. Therefore, the *Leishmania donovani* complex which includes the species *L. donovani*, *L. infantum* and *L. chagasi* was considered because it is the cause for visceral leishmaniasis (Sundar Rai, 2002). Visceral leishmaniasis is the most severe form of leishmaniasis among others (Das *et al.*, 2016). But *tryR* gene sequence of *L. chagasi* was not available in the databases, GenBank, RefSeq, EMBL-EBI or DDJB and therefore the sequences of other two species were used. For HBV, a fully conserved residue longer than 60 nucleotides common to all genotypes was not identified but a partially conserved sequence was selected (Supplementary Table 4).

Designing of potential TALEN target sites

Potential TALEN target sites were identified in each of the above selected conserved residues of the *LMP2A* and *EBNA1* gene of EBV, *E2* gene of HPV type 16, and *tryR* gene of *L. donovani* and *L. infantum*. The tool TALEN Targeter[†] provides both TALEN targets and the RVD sequence for the targets, as the output. Apart from this, it shows the unique restriction sites at the spacer region. Supplementary Table 5 contains the TALEN target results obtained for each sequence. For the *HBx* gene of HBV, the targets were identified for each genotype separately using the whole *HBx* gene as the query, and the targets were identified for a selected, partially conserved sequence, considering all genotypes. Then for each gene, several TALEN targets with higher percentage of HD or NH in their RVDs and having at least one unique restriction site at the spacer region were selected. TALENs having high percentage of HD or NH were selected because the binding specificity and the efficiency are higher when the percentage of HD or NH is high (Streubel *et al.*, 2012). The selection of target sites with unique restriction sites is beneficial in experimental identification of TALEN activity (Doyle *et al.*, 2012).

Potential off-target effect of selected TALENs on human genome, murine genome and unnecessary areas of the pathogen genome

The off-target effect of the TALENs to their targets on human genome, murine genome and unnecessary loci of same pathogen genome were identified for *LMP2A* and *EBNA1* gene of EBV, *E2* gene of HPV type 16 and *tryR* gene of *L. donovani* and *L. infantum*, using the tools PROGNOS and Paired Target Finder. The off-target effect of the TALENs separately selected for the *HBx* gene of each genotype and the TALENs common to *HBx* gene of all the HBV genotypes were also identified.

The off-target effect on the human genome is essential to be identified because if any off-target is present for a 'TALEN respective to the identified target site', it may cause mutations in the human genome. The off-target effect on the mouse genome was also identified as preliminary toxicity tests were mostly carried out *in vivo* using mouse as the model organism, and as such, unnecessary mutations in the mouse genome was avoided. Two tools, Paired Target Finder and PROGNOS, were used in order to minimise the errors in identification of off-target effect. The off-target effects on human and murine genomes were identified with respect to the genomes already available in the tools. These genomes are consensus sequences and therefore, it cannot be concluded that the output given by the tools are valid for every human and mouse, but they would be valid for most. The use of the two tools minimises this effect because the genome entries are different in the two tools.

Other than these, unnecessary bindings on the same pathogen genome were identified in order to minimise the effect of mutations of other genes of the pathogen. If any undesired mutation occurs it might not be suitable for the host or might have a chance to elevate the pathogenic effect. Thus, the tool Paired Target Finder was used because other options were not available in the tool PROGNOS to check the off-targets in NCBI sequences other than the genomes already entered into the tool. Here a selected representative genome was used for each pathogen for convenience and it cannot be concluded about the off-target effect in every isolate and strain of the pathogen.

The score given by the tool, Paired Target Finder is the key by which the tool differentiates the off-targets from the on-targets. The score is given to the TALEN target based on the types of the RVDs present and the matching percentage of RVDs. The perfectly matching off-target gives the same score as the on-target of query TALEN sequence and the score increases when the mismatches increase. Doyle *et al.* (2012) suggested that the maximum score that an off-target would have is four times the score of the on-target. In the tool, only the off-targets below the maximum value are displayed. In the tool PROGNOS, the score has been adjusted to reduce when the off-target deviate from the on-target. Therefore, the perfect off-target is having the score same as that of the on-target. The results of the tool display the off-targets up to a selected number of mismatches, and the off-targets with higher potential of binding with the TALEN are mentioned. The maximum number of mismatches that off-target should contain is selected as five for all the TALEN target sites of the selected genes, and therefore, the off-target effect of those TALENs could be compared.

Identification of off-target effect of the TALENs on genomes of other organisms

The identification of the off-target effect of the TALENs on the genomes of other organisms is necessary. If the designed TALEN proteins are released to the environment, there is a chance of mutating the genomes of other organisms in the environment. The selection of target sites that are lacking off-targets in genomes of other organisms prevents this undesirable effect. Furthermore, human and mouse are inhabited with numerous species of commensals and mutation induction on them can also be predictively prevented with this step. Three methods of BLAST search gave desired results with comparative merits and demerits, and the results were in different formats (Figure 1). In this way the off-target effect of the selected TALENs of *LMP2A* and *EBNA1* of EBV, *E2* gene of HPV type 16, *tryR* gene of *L. donovani* and *L. infantum*, *HBx* gene of all the genotypes of HBV and TALENs common for *HBx* gene of all the genotypes of HBV were identified. The 'nucleotide' search page of the tool BLAST was used because the tools Paired Target Finder and PROGNOS identify only the off-targets in selected genomes. But in BLAST tool it was a challenge to identify the targets/off-targets because the query (the TALEN target site) contained the spacer region which does not involve in the specific binding with a TALEN, and therefore three BLAST search methods were used. The method one of BLAST search displayed the highest number of probable targets/off-targets when compared with other two methods. But the identity and the query coverage the output have been calculated, including the spacer region, although it was masked in BLAST search. This interferes with the differentiation of off-targets from on-targets. In the result type 1 of method two, only the targets in the range one and range two that were lying in a distance not more than 30 nucleotides were selected because the maximum spacer length that a TALEN can be bound is thirty (30) nucleotides (Doyle *et al.*, 2012). In the result type 2 of method two, the results are much effective because the spacer region has not been considered in calculating the percent identity and query coverage. In method three, the results of the on-targets/off-targets are mostly similar to that of the 'result type 2 of method two. In these two results (result type 2 of method two and method three), the identification of the on-targets/off-targets are comparatively easier than other two, referring to the graphical alignment. The suspected sequences for having off-target effects were further tested with the tool Paired Target Finder, because BLAST tool is not specific for the purpose and a TALEN binding score is unavailable in the BLAST tool. Only the sequences

similar to the query sequence (TALEN target sequence) can be identified with BLAST search, and that does not reflect that the TALEN of that query sequence binds with the off-target sequence.

Putative functions of the mutated protein by TALEN

The putative functions of the mutated protein due to the double stranded break by each TALEN of the selected genes were identified. The TALENs that were shown to produce proteins with unnecessary function were avoided. The identification of the putative function of the mutated protein is important because the mutated proteins might have undesirable functions. The proteins that formed due to frame shift of one base pair and frame shift of two base pairs were considered, but frame shift of three base pairs was not considered because it will not change the reading frame and only lead to alterations of a few amino acids resulting slight deviations from the initial function of the protein. The ORFs of the mutated protein that were passing through the cut site of the TALEN were selected because other ORFs are not mutated. Other than that, only the ORFs leading to proteins with amino acid number greater than 75 were considered, because smaller proteins less than that might not possess specific functions. But there are small proteins with key functions in the cell (Reichman-Fried Raz, 2014), and at some point, mutations cause changes in the protein function and is a limitation of the present study. The putative functions of the proteins of the selected ORFs were identified with the protein BLAST tool, where the functions of similar proteins to the query were considered as the function of the query protein. Most of the results obtained for the ORFs were lacking BLAST queries and hence the putative function could not be determined. This might be a potential error because these queries also may have some undesirable functions. This can be identified up to some extent by *de novo* protein function analysis. The partial proteins of the original were not further considered because they might not have novel functions. But some partial proteins could have undesirable functions in the cell.

Selected TALEN target sites

The target sites were selected for each gene that did not show an off-target effect in the human genome and murine genome in both results of the tools Paired Target Finder and PROGNOS, which did not show an unnecessary binding effect in the same pathogen genome and did not produce mutated proteins with undesirable function. Table 1 shows the target sites selected for the *LMP2A* and *EBNA1* genes of EBV, E2 gene of HPV type 16, *tryR* gene of *L. donovani* and *L. infantum* and HBx gene of HBV.

Identification of potential CRISPR/Cas 9 targets

The results obtained for the *LMP2A* gene of EBV is shown in Figure 2. Among them T1, T2, T3, T4, T5,

T6 and T7 were observed to be lacking off targets in the human genome (Supplementary Table 6). The same target sites were obtained and were numbered in the order of off-target effect on the murine genome as T65, T5, T21, T9, T30, T46 and T60 respectively. Among them T5 and T9 were not shown off target effect on the murine genome.

The selected targets were searched for the presence of off-target effect on the genomes of other organisms using 'nucleotide BLAST' tool. From the results it was identified that the target CRLMP2A01 possesses off-target effect on *Schistosoma rodhaini* genome. Therefore, target CRLMP2A02 was selected as the potential CRISPR/Cas9 target on *LMP2A* gene. Any genome with off-target effect with respect to the target CRLMP2A02 was not identified. In the same way CRISPR/Cas9 targets were selected and their potential of binding on other genomes was observed. But targets without off-target effect on human and murine genome were not observed for E2 gene of HPV type 16 and HBx gene of HBV. Table 2 shows the CRISPR/Cas9 targets selected for *EBNA1* gene of EBV, *UL21* and *UL30* gene of HSV-2 and *tryR* gene of *L. donovani* and *L. infantum* respectively. The efficacy of guideRNA of each selected target was analysed by the tool CRISPRater (Labuhn *et al.*, 2018). Score 0.56 or below shows low efficacy, score within 0.56 and 0.74 shows medium efficacy and high

These targets are same as T2 and T4 arrangement s of the off-targets in the human genome.

T2 target of human genome - 5'ACTTGGGATTGCAACACGACGGG3'	}	TARGET CRLMP2A01
T5 target of murine genome - 5'ACTTGGGATTGCAACACGACGGG3'		
T4 target of human genome - 5'TCACGTTTCCTCATCGTTCGGTGG3'	}	TARGET CRLMP2A02
T9 target of murine genome - 5'TCACGTTTCCTCATCGTTCGGTGG3'

Figure 1: An example of the results obtained for the BLAST search for identification of off-targets in other genomes. A: 'BLAST' method one search result type of target site EBVLMP2A04. B: 'BLAST' method two search result type of target site EBVLMP2A05. C: 'BLAST' method three search result type of EBVLMP2A04 target site.

Table 1: Selected TALEN target sites for *LMP2A* and *EBNA1* genes of EBV, (No. 1) *E2* gene of HPV type 16, (No. 2) *tryR* gene of *Leishmania donovani* and *L. infantum* and *HBx* gene of HBV. 'Plus Strand Sequence' is the target sequence as denoted in the plus strand (5'–3'). (No. 3) denotes the TALEN targets separately in each genotype of HBV and (No. 5) denote the targets common for all the genotypes. Supplementary Figure 2 shows the amino acid sequence of TAL01 of target EBVLMP2A01

No.	Pathogen	Gene	Sequence Name	Plus strand sequence
1	EBV	<i>LMP2A</i>	EBVLMP2A01	T GGGGTCCCTAGAAATGGTGC caatgggcgcgggtcc CCCTAGCCCCGCGCGGG A
			EBVLMP2A03	T GGGGTCCCTAGAAATGGTGC ccaatgggcgcgggt CCCCTAGCCCCGCGCGGG A
			EBVLMP2A05	T GGGGTCCCTAGAAATGGTGC caatgggcgcgggtc CCCCTAGCCCCGCGCGGG A
			EBVLMP2A06	T GCCAATGGGCGCGGGTCCCC ctagccccgcggggat CCGGATGGGTACGATGGCGG A
		<i>EBNA1</i>	EBVEBNA01	T GGAGGGGCGAGAGTCTGCA ctccctgtattcaactgagcgtcgg GGGCTGTTGGAGGGGCGAGG A
			EBVEBNA09	T GCACTCCCTGTATTCACTGA gcgtgggggctgtt GGAGGGGCGAGAGTCTGC A
		<i>E2</i>	HPVE201	T TACAAGGCCAGAGAAATGGG atttaaacatattaaccaccagtg GTGCCAACACTGGCTGTATC A
			HPVE202	T ACAAGGCCAGAGAAATGGGA tttaaacatattaaccaccag GTGGTGCCAACACTGGCTGT A
			HPVE203	T AACTGCACCAACAGGATGTA taaaaaacatggatata CAGTGAAGTGCAGTTTG A
			HPVE204	T AACTGCACCAACAGGATGTA taaaaaacatggatata GTGGAAGTGCAGTTTGATGG A
			HPVE205	T TCATGCGGGTGGTCAG gtaatatattgtcct ACATCTGTGTTTAGC A
			HPVE207	T TAACAGCTCACACAAAGGAC ggattaactgtaata GTAACACTACACCCATAGT A
			HPVE208	T AACAGCTCACACAAAGGACG gattaactgtaatag TAACACTACACCCATAGTAC A
			HPVE209	T AACAGCTCACACAAAGGACG gattaactgtaatag AACACTACACCCATAGTAC A
			HPVE210	T TGGACAGGACATAATGTAA acataaaagtgaat TGTTACACTTACATATGAT A
3	<i>Leishmania</i> spp.	<i>tryR</i>	LDTRYR02	T CCCGCGCGTACGACCTCGTG gtgcttgccgcggat CTGGAGGTCTGGAGCGGG A
			LDTRYR03	T GCTTGGCGCCGATCTGGAG gtctggaggcgggatg GAACGCGCCGTCACGCACA A
			LDTRYR04	T GTCCCGCGCGTACGACCTC gtggtgcttgccgc GGATCTGGAGGTCTGG A
			LDTRYR05	T CGGCGGCACGTGCGTGAACG tcggctgcgtgcaaaa GAAACTCATGGTGAC A
			LDTRYR06	T CGGCGGCACGTGCGT gaacgtcggctgcgt GCCAAGAAACTCATGGTG A
			LDTRYR07	T CCGTGAGTCTGGCGG ctteggatgggagatgg ACCGGAATCGCTCTGCCCC A
			LDTRYR08	T TCGGATGGGAGATGGACCG gaatcgtctgcccgaact GGAAGACGCTCATCGCCGCG A
			LDTRYR09	T GGACCGGAATCGCTCTGCC ccaactggaagacgtca TCGCCGGAAGAACA A
4	HBV genotype A	<i>HBx</i>	HBVXGA01	T GGGACTCTCTCGTCCCTTC tcegtctgcgttcca GCCGACCACGGGGCGC A
			HBVXGA02	T TGTCTACGTCCCGTCGGCGC tgaatcccgcggagac CCCTCTCGGGGCCGCTTGGG A
	HBV genotype B	<i>HBx</i>	HBVXGB01	T GGGGCTCTACCGCCCTTC tcegtctgtgtacc GTCCGACCACGGGGCGC A
			HBVXGB02	T TGGGGCTCTACCGCCCGC ttctcegcctgtgt ACCGTCCGACCACGGGGCGC A
	HBV genotype C	<i>HBx</i>	HBVXGC01	T CTACGTCCCGTCGGCGCTGA atccagcggagacc CGTCTCGGGGCCGCTTGGGG A
			HBVXGC02	T CTACGTCCCGTCGGCGCTGA atccagcggagacccc GTCTCGGGGCCGCTTGGGG A
	HBV genotype D	<i>HBx</i>	HBVXGD02	T GCCAAGTGGTCTGCGCGG gacgtcctttgttta CGTCCCGTCGGCGCTGA A
	HBV genotype E	<i>HBx</i>	HBVXGE01	T CCGGCCGACCACGGGGCGCA cctctctttacgcggtct CCCGTCTGTGCCTTCTC A
	HBV genotype F	<i>HBx</i>	HBVXGF02	T GCCGGACCGTGTGCACTTCG ctteacctctgcacgt CGCATGGAGACCACGTGA A
	HBV genotype G	<i>HBx</i>	HBVXGG01	T GGGGCTCTGTCGCCCCCTTC tcegtctgcgttctt GCCGACCACGGGGCGC A
	HBV genotype H	<i>HBx</i>	HBVXGH01	T CGCTTGGGGCTATGCCGCC tcttccgcctgcccgt CCGGCCGACGACGGGTGCG A
			HBVXGH02	T CTCGTGGTCTGCTTGGGGCT atgcccctcttct CCGCTGCCGTTCGGGCCG A
5	HBV	<i>HBx</i>	HBVXCS01	T AGGAGGCTGTAGGCATAAA ttggtctgttacc GCACCATGCAACTTTTTC A
			HBVXCS02	T AGGAGGCTGTAGGCATAA attggtctgttacc AGCACCATGCAACTTTTTC A
			HBVXCS03	T AAATTGGTCTGTTCACCAGC accatgcaactttt CACCTCTGCCTAATC A
			HBVXCS04	T AGGAGGCTGTAGGCATA aattggtctgttacc CAGCACCATGCAACTTTTTC A

Figure 2: Graphical representation of the CRISPR/Cas9 target sites on *LMP2A* gene. A: target sites that arranged according to the increasing order of off-target effect on the human genome. B: target sites that arranged according to the increasing order of off-target effect on the murine genome.

efficacy is shown by scores equal or above 0.74. This scoring system is easily applicable for selecting efficient targets due to its ability to select efficient guideRNA. Supplementary Figure 1 represents the structure and action of CRISPR nuclease.

The protospacer adjacent motif (PAM) selected for the CRISPR/Cas9 targets in this study was NGG of *Streptococcus pyogenes*. But other PAM motifs can be substituted instead of NGG, which might change the off-target effect. The off-targets of the CRISPR/Cas9 nucleases might differ by two base pairs from an on-target (Cho *et al.*, 2014). Therefore, in our study we considered the targets below three pair mismatches as off-targets. The targets above four base pair mismatches mostly prevented the double-strand break and therefore, they were omitted. The cleavage efficiency of CRISPR system greatly varies on different target sites or the cell type/organism. The efficacy of binding and cleavage of CRISPR system depends on several features; features of the guide RNA, genetic features including epigenetics and energetic properties that have been identified through various studies as factors involved in determining the efficacy of the guideRNA (Cong *et al.*, 2013; Fu *et al.*, 2013; Wang *et al.*, 2014; Chari *et al.*, 2015; Liu *et al.*, 2020). The putative function of the mutated protein by CRISPR/Cas9 was not identified because the cut site cannot be exactly determined.

In this research, only the pathogens with double stranded DNA genomes were considered because the TALEN and CRISPR/Cas9 nucleases cannot function on single stranded genomes or in RNA genomes. But recently Abudayyeh *et al.* (2017) have identified the CRISPR/

Cas13 system, which can be applied on RNA genomes.

Delivery of the CRISPR/cas9 nucleases or the TALENs to the required specific cell type is a question not solved yet. Specific strategies should be used for the successful transportation of mRNAs of the nucleases to the cytoplasm and the resulting nucleases to the nucleus of the cell. Viral vectors, microinjection, electroporation and chemical methods are a few currently used methods (Glass *et al.*, 2018). The selection of a specific cell type by the nuclease is also important and Cheng *et al.* (2020) have described a tissue specific nanoparticle based method to deliver CRISPR mRNA.

The development of site specific nucleases is a concerns in terms of an ethical perspective (Rodriguez, 2016). The main ethical concern is the balance between risks and benefits. The loss of ecological equilibrium could occur. Apart from that the regulation of the product to the consumers is also an essential criterion to be evaluated. These ethical questions should be addressed in the development of TALEN and CRISPR/Cas nucleases as treatment strategies.

CONCLUSION

The identified potential TALEN and CRISPR/Cas9 targets may be applicable for specific mutagenic agents of EBV, HPV type 16, HSV-2, HBV and *L. donovani* and *L. infantum* and can be further developed as a treatment strategy. Furthermore, fully conserved residues of enough length for a TALEN target site are absent in the *HBx* gene considering all the genotypes of HBV, and in *E2* gene, considering all high-risk types of HPV.

Table 2: The CRISPR/Cas9 targets selected for the gene of *EBNA1* gene of EBV, *UL21* and *UL30* genes of HSV-2 and *tryR* gene of *Leishmania donovani* and *L. infantum*. (* - targets with high efficacy of mutagenesis)

Pathogen	Gene	Target name	Selected Target	Efficacy score of sgRNA
EBV	<i>EBNA1</i>	CREBNA101	5'GTCGCCGGTGTGTTTCGTATATGG3'	0.66
HSV-2	<i>UL21</i>	CRUL2101	5'TTATTCGTAGGGCGGCCTCGGG3'	0.76*
		CRUL2102	5'GAGCTCGCTTCGTACGTAGTTGG3'	0.72
		CRUL2103	5'GTACGTAGTTGGCGACCATGCGG3'	0.68
		CRUL2107	5'CGCGCCCGACGAACCGACGTTGG3'	0.71
		CRUL2108	5'TCGGCCAACGTCGGTTCGTGCGGG3'	0.56
		CRUL2109	5'CAACGTCGGTTCGTGCGGCGCGG3'	0.81*
		CRUL2111	5'CGTCGGGCGCGGGAACGTACAGG3'	0.81*
		CRUL2113	5'ACTCGAACGCCTGACCCCGGG3'	0.71
		CRUL2114	5'CCCGAGCACGTACGTACCTCGG3'	0.69
	<i>UL30</i>	CRUL3001	5'ATTATCGCCCCGCGTTCGTGG3'	0.67
		CRUL3002	5'CGCCCCAAGGTGTAAGTGGG3'	0.61
		CRUL3003	5'GCCGCGTCGCTTGCCTGTGGG3'	0.70
		CRUL3004	5'GCGTCCCTCGTCCAGCGAACGCGG3'	0.57
		CRUL3005	5'CCAGCTCCACGAGCGATTTATGG3'	0.76*
		CRUL3006	5'CGCGTCGCCGTTACGTCTACGG3'	0.75*
		CRUL3007	5'CCGCGTGCCGTAGACGTGAACGG3'	0.73
		CRUL3008	5'AGACGTGAACGGCGACGCGATGG3'	0.67
		CRUL3009	5'CCCCGCGATCAGGAAGTACGAGG3'	0.72
		CRUL3010	5'CCCGCGATCAGGAAGTACGAGGG3'	0.72
		CRUL3011	5'CGTCGAGTTTAACTGCACGCGGG3'	0.85*
		CRUL3012	5'GATCGGCGAGTATTGTGTGCAGG3'	0.83*
		CRUL3013	5'TTCCGCCGTCGCGCGCTGGCGG3'	0.85*
		CRUL3014	5'GCGCATCATCTACGGGGACACGG3'	0.55
<i>L. donovani</i> and <i>L. infantum</i>	<i>tryR</i>	CRTRYR01	5'GTCCCGCGGTACGACCTCGTGG3'	0.67
		CRTRYR03	5'CACCACGAGGTTCGTACGCGCGGG3'	0.60

REFERENCES

- Abudayyeh O.O. *et al.* (11 authors) (2017). RNA targeting with CRISPR–Cas13. *Nature* **550** (7675): 280–284.
DOI: <https://doi.org/10.1038/nature24049>
- Altschup S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–410.
DOI: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Benjamin R., Berges B.K., Solis-Leal A., Igbinedion O., Strong C.L. & Schiller M.R. (2016). TALEN gene editing takes aim on HIV. *Human Genetics* **135**(9): 1059–1070.
DOI: <https://doi.org/10.1007/s00439-016-1678-2>
- Boldogh I., Albrecht T. & Porter D.D. (1996). persistent viral infections. In: *Medical Microbiology* (eds. S. Baron). University of Texas Medical Branch at Galveston, Galveston, USA.
- Burd E.M. (2003). Human papillomavirus and cervical cancer. *Clinical Microbiology Reviews* **16**(1): 1–17.
DOI: <https://doi.org/10.1128/CMR.16.1.1-17.2003>
- Char S.N., Unger-Wallace E., Frame B., Briggs S.A., Main M., Spalding M.H., Vollbrecht E., Wang K. & Yang B. (2015). Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnology Journal* **13**(7): 1002–1010.
DOI: <https://doi.org/10.1111/pbi.12344>
- Chari R., Mali P., Moosburner M. & Church G.M. (2015). Unraveling CRISPR–Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods* **12**(9): 823–826.
DOI: <https://doi.org/10.1038/nmeth.3473>
- Cheng Q., Wei T., Farbiak L., Johnson L.T., Dilliard S.A. & Siegwart D.J. (2020). Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing. *Nature Nanotechnology* **15**(4): 313–320.
DOI: <https://doi.org/10.1038/s41565-020-0669-6>
- Cho S.W., Kim S., Kim Y., Kweon J., Kim H.S., Bae S. & Kim J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* **24**(1): 132–141.

- DOI: <https://doi.org/10.1101/gr.162339.113>
- Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. & Sayers E.W. (2016). GenBank. *Nucleic Acids Research* **44**(D1): D67–D72.
DOI: <https://doi.org/10.1093/nar/gkv1276>
- Cong L. *et al.* (11 authors) (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**(6121): 819–823.
DOI: <https://doi.org/10.1126/science.1231143>
- Das A., Karthick M., Dwivedi S., Banerjee I., Mahapatra T., Srikantiah S. & Chaudhuri I. (2016). Epidemiologic correlates of mortality among symptomatic visceral leishmaniasis cases: findings from situation assessment in high endemic foci in India. *PLoS Neglected Tropical Diseases* **10**(11): 0005150.
DOI: <https://doi.org/10.1371/journal.pntd.0005150>
- Doyle E.L., Booher N.J., Standage D.S., Voytas D.F., Brendel V.P., VanDyk J.K. & Bogdanove A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Research* **40**(W1): W117–W122.
DOI: <https://doi.org/10.1093/nar/gks608>
- Fan Z. *et al.* (11 authors) (2018). A sheep model of cystic fibrosis generated by CRISPR/Cas9 disruption of the CFTR gene. *JCI insight* **3**(19).
DOI: <https://doi.org/10.1172/jci.insight.123529>
- Feng Z. *et al.* (11 authors) (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research* **23**(10): 1229.
DOI: <https://doi.org/10.1038/cr.2013.114>
- Fine E.J., Cradick T.J., Zhao C.L., Lin Y. & Bao G. (2013). An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Research* **42**(6): e42.
DOI: <https://doi.org/10.1093/nar/gkt1326>
- Frangoul H. *et al.* (27 authors) (2021). CRISPR-Cas9 gene editing for sickle cell disease and β -thalassemia. *New England Journal of Medicine* **384**(3): 252–260.
DOI: <https://doi.org/10.1056/NEJMoa2031054>
- Fu Y., Foden J.A., Khayter C., Maeder M.L., Reyon D., Joung J.K. & Sander J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**(9): 822–826.
DOI: <https://doi.org/10.1038/nbt.2623>
- Glass Z., Lee M., Li Y. & Xu Q. (2018). Engineering the delivery system for CRISPR-based genome editing. *Trends in Biotechnology* **36**(2): 173–185.
DOI: <https://doi.org/10.1016/j.tibtech.2017.11.006>
- Gonzales A.P. & Yeh J.R. (2014). Cas9-based genome editing in zebra fish. *Methods in Enzymology* **546**: 377–413.
DOI: <https://doi.org/10.1016/B978-0-12-801185-0.00018-0>
- Han Y., Teng K., Nawaz G., Feng X., Usman B., Wang X., Luo L., Zhao N., Liu Y. & Li R. (2019). Generation of semi-dwarf rice (*Oryza sativa* L.) lines by CRISPR/Cas9-directed mutagenesis of OsGA2ox2 and proteomic analysis of unveiled changes caused by mutations. *3 Biotech* **9**(11): 1–17.
DOI: <https://doi.org/10.1007/s13205-019-1919-x>
- Hutter G., Bodor J., Ledger S., Boyd M., Millington M., Tsie M. & Symonds G. (2015). CCR5 targeted cell therapy for HIV and prevention of viral escape. *Viruses* **7**(8): 4186 – 4203.
DOI: <https://doi.org/10.3390/v7082816>
- Hwang W.Y., Fu Y., Reyon D., Maeder M.L., Tsai S.Q., Sander J.D., Peterson R.T., Yeh J.J. & Joung J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31**(3): 227–229.
- Kelliher T. *et al.* (11 authors) (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**(7639): 105–109.
DOI: <https://doi.org/10.1038/nature20827>
- Labuhn M., Adams F.F., Ng M., Knoess S., Schambach A., Charpentier E.M., Schwarzer A., Mateo J.L., Klusmann J.H. & Heckl D. (2018). Refined sgRNA efficacy prediction improves large-and small-scale CRISPR-Cas9 applications. *Nucleic Acids Research* **46**(3): 1375–1385.
DOI: <https://doi.org/10.1093/nar/gkx1268>
- Le Sage V., Jung M., Alter J.D., Wills E.G., Johnston S.M., Kawaguchi Y., Baines J.D. & Banfield B.W. (2013). The herpes simplex virus 2 UL21 protein is essential for virus propagation. *Journal of Virology* **87**(10): 5904–5915.
DOI: <https://doi.org/10.1128/JVI.03489-12>
- Leykauf K., Kabsch K., Gassler N., Gissmann L., Alonso A. & Schenkel J. (2008). Expression of the HPV11 E2 gene in transgenic mice does not result in alterations of the phenotypic pattern. *Transgenic Research* **17**(1): 1–8.
DOI: <https://doi.org/10.1007/s11248-007-9130-y>
- Li H.L. *et al.* (11 authors) (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Reports* **4**(1): 143–154.
DOI: <https://doi.org/10.1016/j.stemcr.2014.10.013>
- Liu S. *et al.* (11 authors) (2006). Crystal structure of the herpes simplex virus 1 DNA polymerase. *Journal of Biological Chemistry* **281**(26): 18193–18200.
DOI: <https://doi.org/10.1074/jbc.M602414200>
- Liu Z. *et al.* (11 authors) (2017). Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects CD4+ T cells from HIV-1 infection. *Cell and Bioscience* **7**(1): 47.
DOI: <https://doi.org/10.1186/s13578-017-0174-2>
- Liu G., Zhang Y. & Zhang T. (2020). Computational approaches for effective CRISPR guide RNA design and evaluation. *Computational and Structural Biotechnology Journal* **18**: 35–44.
DOI: <https://doi.org/10.1016/j.csbj.2019.11.006>
- Longnecker R. (2000). Epstein-Barr virus latency: LMP2, a regulator or means for Epstein-Barr virus persistence?. *Advances in Cancer Research* **79**: 175 – 200.
DOI: [https://doi.org/10.1016/S0065-230X\(00\)79006-3](https://doi.org/10.1016/S0065-230X(00)79006-3)
- Marchler-Bauer A., *et al.* (11 authors) (2014). CDD: NCBI's conserved domain database. *Nucleic acids research* **43**(D1): D222 – D226.
DOI: <https://doi.org/10.1093/nar/gku1221>
- McBride A.A. (2013). The papillomavirus E2 proteins. *Virology* **445**(1): 57–79.
DOI: <https://doi.org/10.1016/j.virol.2013.06.006>
- Miller J.C., *et al.* (11 authors) (2011). A TALE nuclease

- architecture for efficient genome editing. *Nature Biotechnology* **29**(2): 143–148.
DOI: <https://doi.org/10.1038/nbt.1755>
- National Cancer Institute of USA (2017). HPV and Cancer. Available at <https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-fact-sheet>, Accessed 25 November 2017.
- Nemudryi A.A., Valetdinova K.R., Medvedev S.P. & Zakian S.M. (2014). TALEN and CRISPR/Cas genome editing systems: tools of discovery. *Acta Naturae* **6**(3): 19–40.
DOI: <https://doi.org/10.32607/20758251-2014-6-3-19-40>
- Ochiai H. & Yamamoto T. (2015). Genome editing using zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). In: *Targeted Genome Editing Using Site-Specific Nucleases* (ed. T. Yamamoto), pp3–24. Springer Japan, Hiroshima, Japan.
DOI: https://doi.org/10.1007/978-4-431-55227-7_1
- Okonechnikov K., Golosova O., Fursov M. & UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**(8): 1166–1167.
DOI: <https://doi.org/10.1093/bioinformatics/bts091>
- O'Leary N.A., et al (11 authors) (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**(D1): D733–D745.
DOI: <https://doi.org/10.1093/nar/gkv1189>
- Park C.Y., Kim D.H., Son J.S., Sung J.J., Lee J., Bae S., Kim J.H., Kim D.W. & Kim J.S. (2015). Functional correction of large factor VIII gene chromosomal inversions in hemophilia A patient-derived iPSCs using CRISPR-Cas9. *Cell stem cell* **17**(2): 213–220.
DOI: <https://doi.org/10.1016/j.stem.2015.07.001>
- Paul M.S., Kaur A., Geete A. & Sobhia M.E. (2014). Essential gene identification and drug target prioritization in Leishmania species. *Molecular BioSystems* **10**(5): 1184–1195.
DOI: <https://doi.org/10.1039/C3MB70440H>
- Reichman-Fried M. & Raz E. (2014). Small proteins, big roles: the signaling protein Apela extends the complexity of developmental pathways in the early zebrafish embryo. *BioEssays* **36**(8): 741–745.
DOI: <https://doi.org/10.1002/bies.201400048>
- Rodriguez E. (2016). Ethical issues in genome editing using Crispr/Cas9 system. *Journal of Clinical Research and Bioethics* **7**(266).
DOI: <https://doi.org/10.4172/2155-9627.1000266>
- Rombel I.T., Sykes K.F., Rayner S. & Johnston S.A. (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene* **282**(1): 33–41.
DOI: [https://doi.org/10.1016/S0378-1119\(01\)00819-8](https://doi.org/10.1016/S0378-1119(01)00819-8)
- Sanders C.M. & Stenlund A. (2000). Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex. *Journal of Biological Chemistry* **275**(5): 3522–3534.
DOI: <https://doi.org/10.1074/jbc.275.5.3522>
- Schwank G., et al (11 authors) (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**(6): 653–658.
DOI: <https://doi.org/10.1016/j.stem.2013.11.002>
- Seeger C. & Sohn J.A. (2014). Targeting hepatitis B virus with CRISPR/Cas9. *Molecular Therapy-Nucleic Acids* **3**(12): e216.
DOI: <https://doi.org/10.1038/mtna.2014.68>
- Shankar S., Prasad D., Sanawar R., Das A.V. & Pillai M.R. (2017). TALEN based HPV-E7 editing triggers necrotic cell death in cervical cancer cells. *Scientific Reports* **7**(5500).
DOI: <https://doi.org/10.1038/s41598-017-05696-0>
- Shen L., Hua Y., Fu Y., Li J., Liu Q., Jiao X., Xin G., Wang J., Wang X., Yan C. & Wang K. (2017). Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. *Science China Life Sciences* **60**(5): 506–515.
DOI: <https://doi.org/10.1007/s11427-017-9008-8>
- Sivachandran N., Wang X. & Frappier L. (2012). Functions of the Epstein-Barr virus EBNA1 protein in viral reactivation and lytic infection. *Journal of Virology* **86**(11): 6146 – 6158.
DOI: <https://doi.org/10.1128/JVI.00013-12>
- Soppe J.A. & Lebbink R.J. (2017). Antiviral goes viral: harnessing CRISPR/Cas9 to combat viruses in humans. *Trends in Microbiology* **25**(10): 833–850.
DOI: <https://doi.org/10.1016/j.tim.2017.04.005>
- Stemmer M., Thumberger T., del Sol Keyer M., Wittbrodt J. & Mateo J.L. (2015). CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PloS One*, **10**(4): e0124633.
DOI: <https://doi.org/10.1371/journal.pone.0124633>
- Stothard P.2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**(6): 1102–1104.
DOI: <https://doi.org/10.2144/00286ir01>
- Streubel J., Blücher C., Landgraf A. & Boch J. (2012). TAL effector RVD specificities and efficiencies. *Nature Biotechnology* **30**(7): 593–595.
DOI: <https://doi.org/10.1038/nbt.2304>
- Sundar S. & Rai M. (2002). Laboratory diagnosis of visceral leishmaniasis. *Clinical and Diagnostic Laboratory Immunology* **9**(5): 951–958.
DOI: <https://doi.org/10.1128/CDLI.9.5.951-958.2002>
- Tsuge M., et al (11 authors) (2010). HBx protein is indispensable for development of viraemia in human hepatocyte chimeric mice. *Journal of General Virology* **91**(7): 1854–1864.
DOI: <https://doi.org/10.1099/vir.0.019224-0>
- Usman B., Nawaz G., Zhao N., Liao S., Qin B., Liu F., Liu Y. & Li R. (2021). Programmed editing of rice (*Oryza sativa* L.) OsSPL16 gene using CRISPR/Cas9 improves grain yield by modulating the expression of pyruvate enzymes and cell cycle proteins. *International Journal of Molecular Sciences* **22**(1): 249.
DOI: <https://doi.org/10.3390/ijms22010249>
- Wang T., Wei J.J., Sabatini D.M. & Lander E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**(6166): 80–84.
DOI: <https://doi.org/10.1126/science.1246981>

Xu P., Tong Y., Liu X.Z., Wang T.T., Cheng L., Wang B.Y., Lv X., Huang Y. & Liu D.P. (2015). Both TALENs and CRISPR/Cas9 directly target the HBB IVS2–654 (C>T) mutation in β -thalassemia-derived iPSCs. *Scientific reports* 5(12065).

DOI: <https://doi.org/10.1038/srep12065>

Yeadon J. (2014). Pros and cons of ZFNs, TALENs, and CRISPR/CAS. The Jackson Laboratory Blog Post.

Available at <https://www.jax.org/news-and-insights/jax-blog/2014/march/pros-and-cons-of-znfs-talens-and-crispr-cas>, Accessed 4 November 2017.

Yuen K.S., et al (11 authors) (2015). CRISPR/Cas9-mediated genome editing of Epstein-Barr virus in human cells. *Journal of General Virology* 96(3): 626–636.

DOI: <https://doi.org/10.1099/jgv.0.000012>

September 2021

Supplementary Tables 1-6: Use below link to access Supplementary Tables 1-6
<https://drive.google.com/drive/folders/1k19gtfOQsK6bcmR1EMH2qti2az-PG0tI?usp=sharing>