Research Article
Open Access

CRISPR construct covering homo sapiens APOL (1-6) genes

Abstract

Among the several design tools available for CRISPR/Cas9genome editing, E-CRISP web application was chosen to design single guide RNA (sgRNA) because it provides flexible output, enabling design of multiple libraries. Recently, two coding sequence variants in APOL1, which encodes a trypanolytic factor in humans and gorillas, have been implicated to associate with kidney disease which could result from inaccurate genome editing. There is need to design accurate library of guided RNA for not only APOL1 but also APOL 2-6 for CRISPR/Cas9 genome editing experiment through which one can apply their products in real life experiment.

Method: E-crisp web server was used to design library of single guided RNA (sgRNA) for CRISPR/Cas9 experiments with high specificity, efficacy and annotation covering the Homo sapiens APOL (1-6) genes.

Result: The program was able to design library of single guided RNA (sgRNA) for CRISPR/Cas9 experiments with high specificity, efficacy and annotation covering the Homo sapiens APOL (1-6) genes. APOL6 has highest percentage of library cover with no off-target observed.

Keywords: CRISPR/Cas9, sgRNA, APOL (1-6), library, gene editing

Volume 10 Issue 1 - 2021

Akinseye Olanrewaju Roland,[1] Ojomo Joan L,[1] Ebenezer Morayo Ale,[1] Gbadamosi Folawiyo I[2]

[1]Nigerian Institute of Trypanosomiasis & Onchocerciasis Research, Ibadan, Nigeria
[2]Department of Biochemistry, Faculty of Pure and Applied Sciences, Federal University Wukari, Nigeria

Correspondence: Akinseye Olanrewaju Roland, Nigerian Institute of Trypanosomiasis & Onchocerciasis Research, Ibadan, Nigeria, Email Akinseyeroland@gmail.com

Received: April 18, 2021 | **Published:** June 11, 2021

Abbreviations: CRISPR-Cas9, clustered regularly inter-spaced palindromic repeats-CRISPR associated proteins 9; crRNA, CRISPR RNA; sgRNA, single guide RNA; APOL, apolipoprotein

Introduction

There are numerous microorganisms especially viruses, attacking other microbes which constitute incessant threat; couple with the multiplication and mutation rates observed in viruses, it makes them predators that evolved very fast.

To survive the assault of the invading viruses and predators, bacteria has developed a multilayered defense mechanism against viral DNA known as CRISPR-Cas9.[1] From many tools for genome editing,the most reliable method is CRISPR-Cas9, due to the simple way it form a break in the strand of helix engineer byguided-RNA nucleases which include pairing of base between the target DNA region and the designed RNA that will attached to it. As a results of the modifications in the genome, these provide a feasible opportunities for not only transmitting good traits but also to correct malfunctioned genes for curatives with use in gene therapy.[2]

Every CRISPR encodes an RNA (crRNA), comprises of a guide RNA (gRNA) and CRISPR RNA parts that are trans-activating. There is embedding of a crRNA processed fragment into the Cas9 protein, directing it to the DNA being targeted, position at which a break of double-strand is introduced by the Cas9 nuclease.[3,4] The CRISPR-Cas method has been successfully employed to modify gene function including in human stem cells, mice amidst other organisms. CRISPR-Cas9 system is a useful tool in the biological related research especially to adjust genes via knockout studies, inserting specific sequences, up-regulation and down-regulation of genes.

The broad usefulness of CRISPR through gene editing and the accompanying simple rules for designing sgRNA has furnished an exciting advantage for researchers in the bioinformatics field to have a broad view of diverse aspects of CRISPR mechanism employing computational approach and designing of software which specifies possible sgRNA sequences in genomes.

Among the various software tools and web servers available for gene editing, E-CRISP web server was chosen to design single guide RNA (sgRNA). It delivers adjustable result and designed parameters that are practically patterned, providing opportunity for multiple libraries designs and hence organized analysis of the effect of various parameters. E-CRISP identifies target sequences complementary to the gRNA ending in a 3′ protospacer-adjacent motif (PAM), N(G or A)G, which is required for the recruited Cas9 nuclease to cut the DNA double strand. The task of E-CRISP is to find CRISPR target sites which are specific only for the sequence enter by the user, and no other sequence. Bowtie2 is employed in e-crisp to test for the specificity by mapping the identified gRNA sequence to the rest of the organisms chromosomal DNA (default), gene models or mRNA. If it can map the gRNA to another sequence, it has an off-target.

The family of gene known as apolipo protein L (APOL) which composed of six genes (APOL 1-6) as shown in Figure 1, in humans, which are grouped within 619 kb on human chromosome 22.[5] The APOL1 is the most studied family member,in which a trypanolytic protein is encodedin humans and gorillas, killing pathogenic Trypanosoma brucei subspecies during bloodstream infections.[6]

Recently, two coding sequence variants in APOL1 have been shown to associate with kidney disease in a recessive fashion while at the same time conferring resistance against Trypanosoma brucei rhodesiense.[7] The presence of the C-terminal polymorphisms cause a resistance which is due, in part, to decreased binding of the G1 and G2 APOL1 variants to the T. b. rhodesiense virulence factor, serum resistance-associated protein (SRA).[8] People who have at least one copy of either the G1 or G2 variant are resistant to infection by trypanosomes, but people who have two copies of either variant are at

an increased risk of developing a non-diabetic kidney disease. There is need to design library of guided RNA for not only APOL1 but also APOL 2-6 for CRISPR/Cas9 genome editing experiment through which we can studied their products in real life experiment.

This study design library of single guided RNA (sgRNA) for CRISPR/Cas9 experiments with high specificity, efficacy and annotated covering the Homo sapiens APOL (1-6) genes which is a preliminary step before confirmatory in vitro/vivo analysis is carried out.

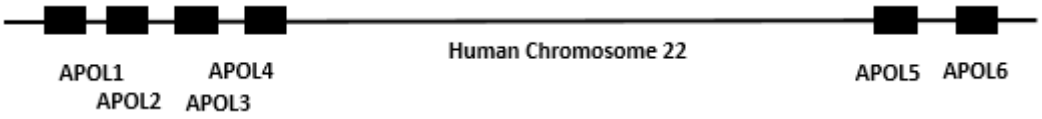

Figure 1 Genomic organization of the six human APOL genes.

Materials and methods

The method describe by Michaela et al., 2014 was used for the design.[9] The first section begins by chosen the organism the design should be made for. The databases of ECRISP are initially built for each organism. In this case, Homo sapiens GRCh 38 was chosen and followed by the subsequent input of the genes symbol APOL 1-6 which CRISPR will be designed to target.

The purpose of design the editing experiment was specify as knockout option because only coding sequences of the APOL 1-6 was targeted and the need for the designs to take care of the downstream of the start codon.3' protospacer adjacency matrix(PAM) IUPAC base pair code was set to NGG while all other options were left as default. In gene annotation filtering section, all results which do not target anexon were excluded from the output.

The alignment program Bowtie2 in e-crisp was used to evaluate the Off-target effects and target-site homology. Chromosomal DNA was selected as database and the program should search for secondary off-target. The off-targets are analyzed on the basis of sequence alignment of each design to the reference genome. The output section was specify to create an image showing genomic context, produce result table to the browser window and produce additional information for the match string.

Results and discussion

CRISPR design covering APOL (1-6) genes

CRISPRs were predicted for 6 genes with a predicted apolipo protein domain in Homo sapiens. Off-target analysis was performed using complete genome sequence as database (genome Homo sapiens GRCh38). In reporting valid on-target seen in Figure 2A (which was defined as perfect match with no mismatch), designs with more than 1 valid target were excluded while in figure 2B, a valid off target was allowed to contain more than 1 mismatches. In consequence, when allowing for imperfect matches in target (and off-target) sites, a higher number of designs has more off-targets and thus was excluded. CRISPR target sequence length was 23 bp comprising a 20 bp protospacer and the 3 bp PAM (NGG). Only the 17 bp counted from the 3' end were considered for specificity testing. Designs mapping outside of exons were excluded.

The number of hits is the number of locations the CRISPR design targets or number of times it appear in the output table. The criterion for selection of the best sgRNA target sequence for each gene was to choose coding exon with number of hit equals to 1. The library cover per gene was calculated by dividing the number of hit equals to 1 for each gene with the total number of successful designs. The Y-axis is normalized to the percentage of genes in the library that is covered by the criterion. The criterion (Xaxis) is the designs per gene coverage. No designs targeting introns were allowed. The number of hits began with 2 in Apol 5, hence no unique design was recorded for it (Figure 2).

APOL (1-6) genesg RNA library for homo sapien

CRISPR target sites have been predicted for every gene that is annotated as an apolipoprotein in Homo sapien. The database used in off-target analysis was genomic DNA (genome Homo sapiens GRCh38).The 6 most 5' bases were not considered for specificity assessment (Table 1).

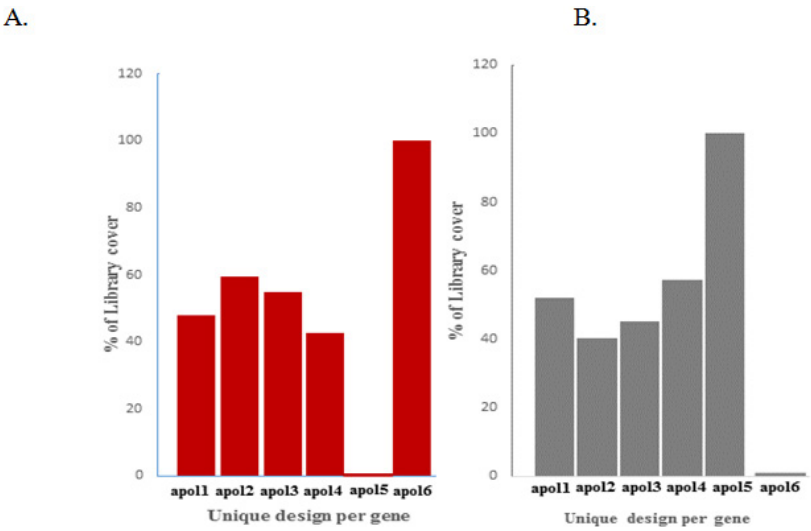

Figure 2 (a-b): Percentage of library cover by CRISPR design per apol gene (1-6); Unique design per gene – apol1 (49), apol2 (31), apol3 (39), apol4 (23), apol5 (0) and apol6 (14).

Citation: Roland AO, Joan OL, Ale EM, et al. CRISPR construct covering homo sapiens APOL (1-6) genes. *MOJ Proteomics Bioinform.* 2021;10(1):9–11. DOI: 10.15406/mojpb.2021.10.00292

Table 1 APOL (1-6) genes gRNA library for Homo sapien

Gene symbol	Ensembl identifier	gRNA target sequence	PAM
APOL1	ENSG00000100342	GCAGGGCCTCCTCCTTGAGC	TGG
APOL2	ENSG00000128335	GCCGCTGGCACCATGAACCC	TGG
APOL3	ENSG00000128284	GCAGATGCACGGCTGGAGGT	CGG
APOL4	ENSG00000100336	GTGCAGCTCATCACAAGCGT	GGG
APOL5	ENSG00000277390	GTCTGTGCCCCAAGGATGCT	TGG
APOL6	ENSG00000221963	GAGGCTGATGGACAACCAGG	CGG

CRISPR target sequence length was 23 bp comprising of a 20 bp protospacer and a 3bp protospacer adjacency matrix PAM (NGG).Only the 17 bp counted from the 3' end were considered for specificity testing. Designs mapping outside of exons and with more than 1 mapped location were excluded. Valid alignments were allowed to have no mismatch. CRISPR target sequence length was 23 bp comprising a 20 bp protospacer and a 3bp protospacer adjacency matrix (NGG). The best scoring design, preferably targeting the first coding exon is shown for each gene in Table 2.

Table 2 Specificity Score, Annotation score and Efficacy score of APOL 1-6

Gene symbol	S-score	A-score	E-score
APOL1	100	32.1522	62.0067
APOL2	100	25.2192	68.8636
APOL3	100	63.8605	65.5205
APOL4	100	40.4347	76.6198
APOL5	84.3	37.5	52.5343
APOL6	100	58.3333	82.6685

Specificity score (S-score); annotation score (A-score); efficacy score (E-score)

In the scoring of sgRNA designed for CRISPR/Cas9 experiments, in addition to the usual scores of Doench et al. and Xu et al. which was given in the output table, E-CRIS Pemployed Specificity score (S-score), Annotation score (A-score) and Efficacy score (E-score).

Specificity Score was generated by starting with 100 for every off-target substract (20- mismatches)/iteration.[9] Annotation score was estimated starting from 0 for every hit exon, adding 5 per exon counted and codon sequence hit and addition of 1 to every start/stop codon and gene hit.[9] Efficacy score was computed by addition of 1 if the the last 6 bp have a CG, sequence is preceded by a G, there are GG in front of the target sequence (opposite the PAM) and substract 1 if the entire sequence has GC content greater than 80%.[9]

Conclusion

Since the main task of E-CRISP is to find CRISPR target sites which are specific which only target the APOL 1-6 inputted and no other sequence, the program was able to design library of single guided RNA (sgRNA) for CRISPR/Cas9 experiments with high specificity, efficacy and annotation covering the Homo sapiens APOL (1-6) genes at shortest time. As a results of side effects observed in genomic variants interactions, CRISPR/Cas9 gene editing application should incorporate features to assess stability of sg RNA designed in laboratory experiment in future update.

Acknowledgments

None.

Conflicts of interest

The authors declare that there is no conflict of interest.

References

1. Choi KR, Lee SY. CRISPR technologies for bacterial systems: current achievements and future directions. *Biotechnol Adv*. 2016;34(7):1180–1209.
2. Ragothaman M Yennmalli, Siddhant Kalra, et al. Computational tools and resources for CRISPR/Cas 9 genome editing method. *MOJ Proteomics Bioinform*. 2017;5(4):116–120.
3. Gasiunas G, Barrangou R, Horvath P, et al. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria *Natl Acad Sci.* USA 109. 2012;E2579–E2586.
4. Deltcheva E. Krzysztof Chylinski, Cynthia M Sharma, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011;471:602–607.
5. Page NM, Butlin DJ, Lomthaisong K, et al. The human apolipoprotein L gene cluster: Identification, classification, and sites of distribution. *Genomics*. 2001;74:71–78.
6. Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, et al. Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature*. 2003;422:83–87.
7. Schaub C, Verdi J, Lee P, et al. Cation channel conductance and pH gating of the innate immunity factor APOL1 is governed by pore lining residues within the C-terminal domain. *J Biol Chem*. 2020;295:13138–13149.
8. Fontaine F, Lecordier L, Vanwalleghem G, et al. APOLs with low pH dependence can kill all African trypanosomes. *Nat Microbiol*. 2017;2:1500–1506.
9. Michaela Spitzer, Jan Wildenhain, Juri Rappsilber, et al. E-CRISP: fast CRISPR target site identification. *Nature Method*. 2014;11(2):122.