Genomics Paper

CRISPR-sub: Analysis of DNA substitution mutations caused by CRISPR-Cas9 in human cells

Eslam Abed Abd-elwahab

19 April, 2021

Abstract

CRISPR-Cas9 induces DNA cleavages at desired target sites in a guide RNA-dependent manner; DNA editing occurs through the resulting activity of DNA repair processes including non-homologous end joining (NHEJ), which is dominant in mammalian cells. NHEJ repair frequently causes small insertions and deletions (indels) near DNA cleavage sites but only rarely causes nucleotide substitutions. High-throughput sequencing is the primary means of assessing indel and substitution frequencies in bulk populations of cells in the gene editing field. Here, we developed a novel analysis method, named CRISPR-Sub, to statistically detect Cas9-mediated substitutions in high-throughput sequencing data by comparing Mock- and CRISPR-treated samples. We first pinpointed 'hotspot positions' in target sequences. By examining 51 endogenous target sites in HeLa cells, we found that the average apparent substitution frequency was 0.8% in all queries, that apparent substitutions frequently occur near CRISPR-Cas9 cleavage sites, and that nucleotide conversion showed no meaningful nucleotide preference patterns, strongly suggesting that DNA substitutions are generated by the NHEJ pathway.

Introduction

The CRISPR (clustered regularly interspaced short palindromic repeat)-Cas , an adaptive immune response in bacteria and archaea. DNA double-strand breaks (DSBs) induced by CRISPR endonucleases are typically repaired by a cell's own repair processes, such as the homology-directed repair (HDR) pathway, the non-homologous end joining (NHEJ) pathway, or an alternative KU-independent process such as the microhomology-mediated end joining (MMEJ) pathway . Among these, the error-free HDR pathway, in the presence of an inserting donor DNA, is useful for generating targeted gene knock-ins or gene corrections , whereas the error-prone NHEJ pathway is frequently accompanied with small insertions and deletions (indels) and MMEJ causes DNA sequence-dependent deletions, both resulting in gene disruption at desired target sites. Among various CRISPR effectors type II Cas9 derived from *Streptococcus pyogenes* (SpCas9) is the most widely used due to its high efficacy and simple DNA recognition sequences (5′-NGG-3′), which are also called protospacer adjacent motifs (PAMs). A few studies have reported the presence of single nucleotide substitutions in DNA cleavage sites after CRISPR-Cas9 treatment . However, these substitution mutations caused by CRISPR-Cas9 have not attracted serious attention and are typically thought to be an exception. When gene editing is assessed in such populations in high-throughput manner using next generation sequencing (NGS) technologies, DNA substitutions are normally tainted by substitution errors, which are mainly derived from the DNA amplification process and DNA sequencers . In this study, we developed a novel analysis method, CRISPR-Sub, to measure the frequency of Cas9-mediated

substitutions by comparing high-throughput sequencing data from Mock- and CRISPR-treated samples.

Related works

1) **P.D. Hsu, E.S. Lander, F. Zhang**, Development and Applications of CRISPR-Cas9 for Genome Engineering , refers to the process of making targeted modifications to the genome, its contexts (e.g., epigenetic marks), or its outputs (e.g., transcripts). The ability to do so easily and efficiently in eukaryotic and especially mammalian cells holds immense promise to transform basic science, biotechnology, and medicine .
2) **K. Rothkamm, I. Krüger, L.H. Thompson, M. Löbrich**, Pathways of DNA double-strand break repair during the mammalian cell cycle, Using immunofluorescence detection of γ-H2AX nuclear foci as a novel approach for monitoring the repair of DSBs, we show here that NHEJ-defective hamster cells (CHO mutant V3 cells) have strongly reduced repair in all cell cycle phases after 1 Gy of irradiation.
3) **M.R. Lieber, Y. Ma, U. Pannicke, K. Schwarz** ,Mechanism and regulation of human non-homologous DNA end-joining, Double-stranded DNA breaks are repaired by two main pathways — homologous recombination and non-homologous DNA end-joining (NHEJ). By comparing the use of these two pathways, we can gain an insight into their biological functions in yeast and multicellular eukaryotes.
4) **K.M. Esvelt, P. Mali, J.L. Braff, M. Moosburner, S.J. Yaung, G.M. Church**, Orthogonal Cas9 proteins for RNA-guided gene regulation and editing , Clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated (Cas) systems provide bacteria and archaea with acquired immunity by incorporating fragments of viral or plasmid DNA into CRISPR loci and using the transcribed crRNAs to guide the degradation of homologous sequences.
5) **K. Chylinski, K.S. Makarova, E. Charpentier, E.V. Koonin**, Classification and evolution of type II CRISPR-Cas systems, Two subtypes of type II CRISPR-Cas systems are defined on the basis of their characteristic operon organizations. Type II-A systems encompass an additional gene, known as *csn2*. The Csn2 protein is involved in spacer integration but is not required for interference. Two distantly related Csn2 subfamilies that include short and long forms of the protein have been identified. For both of these forms, the structures have been solved and the proteins have been biochemically characterized.
6) **E.V. Koonin, K.S. Makarova, F. Zhang Diversity**, classification and evolution of CRISPR-Cas system ,Two classes of CRISPR-Cas systems differ by the architectures of effector modules.Effectors of Class 2 CRISPR-Cas systems are large, multidomain proteins.Effector modules of Class 2 CRISPR-Cas systems independently evolved from transposon genes.Some Class 2 CRISPR-Cas effectors are also involved in pre-crRNA processing.

Methods

1. **Generation of sgRNA-encoding plasmids** , Each oligo including sgRNA . Oligos were heated and cooled down to make double-strand oligos.

2. **Cell culture and transfection,** For CRISPR-treated cells, 750 ng of SpCas9 expression plasmid and 250 ng of sgRNA expression plasmid were mixed with 100 µl of Opti-MEM medium and 2 µl of Lipofectamine 2000 and incubated for 20 min at room temperature For Mock-treated cells, the mixture contained 750 ng of SpCas9 expression plasmid only without sgRNA expression plasmid.

3. **LIG4 knockout cell lines**,The presence of *LIG4* mutations in each cell line was determined; cell lines with confirmed mutations were used in later experiments.

4. **Targeted deep sequencing**, The targeted region (200–270 nucleotides) of genomic DNA was amplified using Phusion polymerase. The PCR products were subjected to paired-end read sequencing using Illumina Mini-seq.

5. **Off-line tool for detecting CRISPR-induced substitutions**, extracts the relevant sequence information from the NGS results and removes the sequences with indels by comparing their length to that of the wild-type sequence. From the NGS outcomes, we calculated (Sub-fold) value and a frequency of count at each position.

Results

1. Strategy for calculating the frequency of apparent CRISPR-induced substitutions in a bulk population of cells, we expected to measure the total frequency of bona fide substitutions statistically by comparing CRISPR-treated sample data against Mock-treated sample data as a negative control, although it is almost impossible to distinguish whether any given substitution is CRISPR-derived or a false positive.

2. Analysis of apparent substitutions in the PYK2 gene as a proof-of-concept

3. Comprehensive evaluation of apparent substitutions in 50 additional endogenous human targets

4. The NHEJ pathway is a major cause of the observed nucleotide substitutions

5. CRISPR-induced substitutions in human HEK293T cells