
crisprseeker Documentation

Release 0.99.0

Alper Kucukural, Sophie Wigmore

July 18, 2016

1	What is Compare 2 Sequences?	3
2	What is Off Target Analysis?	5
3	What is GUIDE-seq Analysis?	7
4	Using the Interface	9
5	About Files	11

Contents:

crisprseeker v1.0.0

What is Compare 2 Sequences?

Compare 2 Sequences generates all possible guide RNAs (gRNAs) for two input sequences, or two sets of sequences and generates scores for potential off-targets in the other sequence.

The two input files contain the two sequences to be searched for potential gRNAs. Compare 2 Sequences assigns relative cleavage scores to each gRNA for both input sequences, facilitates identification of gRNAs that specifically target one of the two input sequences, and facilitates identification of gRNAs that target both input sequences.

It returns a data frame with all potential gRNAs from both sequences. In addition, a tab delimited file scores-For2InputSequences.xls is also saved in the outputDir, sorted by scoreDiff descending. Output gRNAs are in fasta or genbank formats, in paired configuration, and with restriction enzyme cut sites.

What is Off Target Analysis?

Off target analysis is the design of target-specific guide RNAs (gRNAs) for CRISPR-Cas9 system by automatically calling findgRNAs, filtergRNAs, searchHits, buildFeatureVectorForScoring, getOfftargetScore, filterOfftarget, calculating gRNA cleavage efficiency and generate reports.

The program finds potential gRNAs for the input sequence, searches and scores off-target sites for each gRNA, and retrieves genomic sequences flanking off-target sites. It then indicates whether the off-target sites are in critical region of the gene such as exon and outputs off-target details including mismatch position, cleavage score, and gRNAs with topN off-target cutting scores.

Four tab delimited files are generated in the output directory: OfftargetAnalysis.xls (detailed information of off targets), Summary.xls (summary of the gRNAs), REcutDetails.xls (restriction enzyme cut sites of each gRNA), and pairedgRNAs.xls (potential paired gRNAs)

What is GUIDE-seq Analysis?

GUIDE-seq analysis workflow includes functions for obtaining unique insertion sites (proxy of cleavage sites), estimating the locations of the insertion sites, aka, peaks, merging estimated insertion sites from plus and minus strand, and performing off target search of the extended regions around insertion sites.

This method relies on erroneous NHEJ-mediated DNA repair to capture co-introduced blunt-ended double stranded oligonucleotides (dsODNs) at Cas9-induced breakpoints within the genome. The GUIDE-seq dsODNs display high insertion frequency (up to 50% of the measured indel rate (Tsai et al., 2015)) at Cas9- induced DSBs, thereby tagging these loci for selective amplification and subsequent deep sequencing. The method is quite sensitive as off-target sites with $>0.1\%$ indel frequency can be detected, and the frequency of dsODN insertion appears to be correlated with the frequency of Cas9-induced lesions at each site (Tsai et al., 2015). This method has been used successfully to evaluate the precision of Cas9 and its variants (tru-sgRNAs (Tsai et al., 2015) or PAM variants (Kleinstiver et al., 2015)). Given its favorable properties, GUIDE-seq could become a standard in the nuclease field for off-target analysis. While the GUIDE-seq method is straightforward to employ, to date no bioinformatic tools have been released to the community to support the analysis of this data.

We developed GUIDESeq package to facilitate the analysis of GUIDE-seq dataset, including retaining one read per unique molecular identifier (UMI), filtering reads lacking integration oligo sequence (dsODNs), identifying peak locations (cleavage sites) and heights, merging cleavage sites from plus strand and those from minus strand, and performing target and off target search of the input gRNA. This analysis leverages our ChIPpeakAnno package (Zhu et al., 2010) for merging cleavage sites from plus strand and minus strand, and CRISPRseek package (Zhu et al., 2014) for defining the homology of any identified off-target site to the guide sequence and Cas9 PAM specificity.

Using the Interface

The user interface on CRISPRSeek allows control over offTargetAnalysis and compare2Sequences by allowing you to change preset defaults depending on your data files.

First, choose which analysis you would like to do, either off target analysis, compare 2 sequences, or GUIDE-seq analysis. Afterwards, upload your files using the file uploader. If you do not have any files to upload, the default demo files will be used in the analysis instead. The default output directory is a temporary directory.

Next, change default settings under the Submissions panel accordingly in the main panel and in the advanced settings panel if needed. Click “Submit” to proceed with analysis, or “Reset all fields to defaults” to reset all the settings you changed.

After analysis is complete, you may download the output data as a zip file by clicking the “Download” button on the bottom left, and selecting where you would like to save it.

Under the Data Table panel, upon completing analyses on the data, sample data tables will be available. For off target analysis, the example data table is for the “RE Cut Details” output. For compare 2 sequences, the shown data table is the “Scores for 2 Input Sequences” output file. Lastly, for GUIDE-seq, the output data table is the data from the “gRNA Peaks” output file.

About Files

The maximum file size that can be uploaded is 1 GB. Anything larger will not be accepted by the application. If no files are uploaded, the demo files will be used in analysis. Adding large files may take a while to upload, as well as analyze. Refer to the prompt at the top of the screen that will tell if the file is still being analyzed, or if the analysis is complete.