OffRisk User guide

OffRisk is a pipeline that comprise of two dockers – server docker and web UI docker.

The main function of the server is to analyze off-target and return the result of the analysis from different databases.

The web UI enables users to easily run analysis with offRisk server for two use cases:

- 1. Off-target analysis input from the user for off-target location and returning the results of the analysis.
- 2. On-target analysis user input gRNA sequences, and then using Cas-OFFinder and FlashFry the pipeline will search for the off-target locations and then analysis them and return the results.

Installation

Docker

To work with the docker files, docker and docker-compose must be installed.

https://docs.docker.com/get-docker/

If downloading for windows use the "WSL 2 backend" option

- Enable the WSL 2 feature on Windows. For detailed instructions, refer to the <u>Microsoft documentation (https://learn.microsoft.com/en-us/windows/wsl/install)</u>.
- Download and install the <u>Linux kernel update package (https://learn.microsoft.com/enus/windows/wsl/install-manual#step-4---download-the-linux-kernel-update-package)</u>.
- Set WSL 2 as your default version, as it is needed by the docker.

Restart computer and run "Docker desktop" from the start menu / desktop shortcut (if you created one)

The dockers configurations are in docker-compose.yml file for easy work with both dockers. A modification to the volumes path needs to be done – please see explanation in chapter Run ontarget / off-target search on docker.

To work with OffRisk the following file need to be downloaded: https://www.ise.bgu.ac.il/clusters/OffRisk-db.tar.gz

This will require 8 GB. It contains the database folder, this user guide and docker-compose.yml file.

The docker are also available to download without the docker-compose.yml file from:

OffRisk dockers are on DockerHub: https://hub.docker.com/r/almaliahbgu/off risk/tags

- Docker server will require 6.5 GB.
- Docker UI will require 1.61 GB.

The code is also available from GitHub in the following links:

- https://github.com/gili311/OffRisk

- https://github.com/gili311/OffRisk-ui

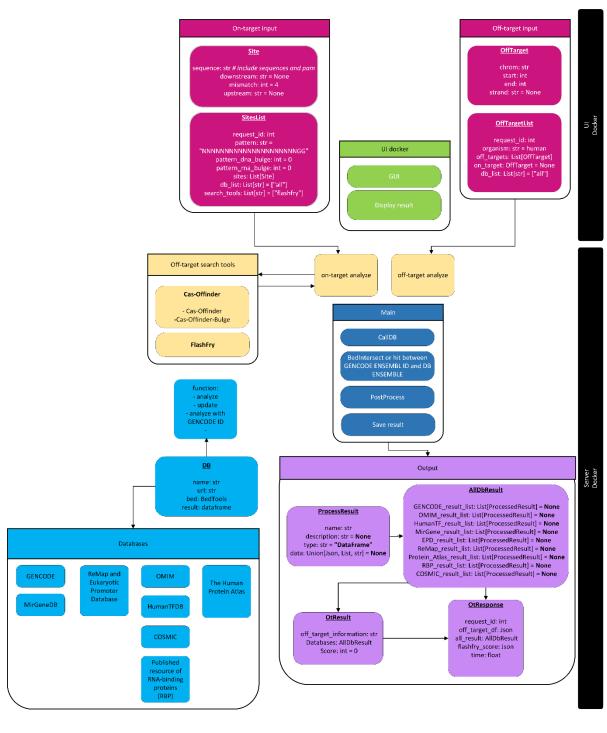
Usage

The web UI docker must communicate with the server docker.

The server docker can work as standalone and integrate with other tools by using API.

Validation on the input is done using python module pydantic and have the following structure:

General structure



Off-target

Has two types of objects:

```
OffTarget():
    chrom: str
    start: int
    end: int
    strand: str = None

OffTargetList:
    request_id: int
    organism: str = 'human'
    off_targets: List[OffTarget]
    on_target = OffTarget = None
    db list: List[str] = ["all"]
```

OffTarget defines one off-target. "chrom" structure begins is a number of the chromosome or "X", "Y". "strand can be "+" or "-".

OffTargetList is an object containing request id for API request, off_sites which is a list of OffTarget and db list which contains specification on which databases to run the analysis.

On-target

Has two types of objects:

Site defines how a single gRNA we are searching for will look like. The sequence of the gRNA is mandatory and must include the PAM site, and mismatch is the number of allowed mismatches. By default, it will be 4.

SiteList is an object containing request id for API request, a list of sites to search for, db_list which contains specification on which databases to run the analysis. Search_tools it the tool to search for off-target and can be FlashFry or Cas-OFFinder. Finally, the different pattern options are as described in Cas-OFFinder documentation.

Database folder structure

Database folder structure:

```
databases
   COSMIC
    └─ cosmic.csv
   EnhancerAtlas

    enhanceratlas.bed

   FlashFry
      - cas9ngg_database
       · cas9ngg database.header
   GENCODE
    gencode.v42.chr patch hapl scaff.annotation sort.gff3
    genome
     — hg38.fa
   HumanTFDB
    — human tf.csv
   MirGeneDB
    MIMO
    — omim.csv
    Pfam protein domains
    pfam protein domains.bed
   ProteinAtlas
      - protein atlas.csv
    RBP
    └─ rbp.csv
   ReMap_EPD

    remap epd.bed

    TargetScan
      · targetscan.bed
```

"db list" contains a list of supported databases to analyze.

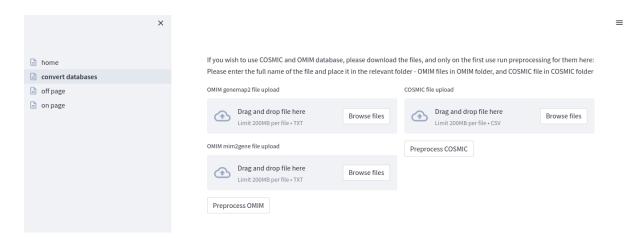
Currently supported databases:

```
"gencode", "mirgene", "remapepd", "enhanceratlas", "pfam", "targetscan", "omim", "humantf", "protein atlas", "rbp", "cosmic"]
```

All relevant files for the databases are in the OffRisk-db.tar file except for OMIM and COSMIC. OMIM and COSMIC require licenses. therefore, users who want to use them need to download them from the website and convert them:

- Download the file for COSMIC: from https://cancer.sanger.ac.uk/cosmic/download, file Cancer Gene Census.
- 2. Download the files for OMIM. Both files are needed: from https://www.omim.org/contact , files name: mim2gene.txt and genemap2.txt.
- 3. Navigate in OffRisk to the page **convert databases**, and upload the files there.
- 4. For COSMIC database conversion click on the button **Pre-process COSMIC**.
- 5. For OMIM database conversion click on the button **Pre-process OMIM**.
- 6. Download the result file (do not change the given name) and place it in the relevant folder on your database folders.

- 7. The file for COSMIC should be **cosmic.csv** in folder **COSMIC**.
- 8. The file for OMIM should be **omim.csv** in folder **OMIM**.



A script name preprocess.py is also available for converting new files and versions of the original database to the structure of OffRisk databases. The script should be run with name as

Run on-target / off-target search on docker

on the first use run the following:

- 1. Change the path of the volume for your database in docker-compose.yml under services>off-risk-server->volumes.
 - For example, if the path of the database is in /home/database, then the line should be: "/home/database:/databases"
- 2. Run: 'docker network create OffRisk-net'.
- 3. Run: 'docker-compuse up -no-build'

once both docker are up browse to the one of the URL presented like the following example (Network URL or External URL:

```
off-risk-ui_1
```

or to the following default URL: http://localhost:8051/.

In the UI side bar, there are 4 options under Navigation:

- Home home page.
- Convert databases for creating OMIM and COSMIC databases.
- Off-target page to analyze input off-target sites from the user.
- On-target page to search for off-target for a gRNA and then analyze the resulted locations.

Once choosing the relevant page follow the instruction:

Off-target page

Select the relevant server and database for this analysis. Options are CRISPR-IL, local or custom. For local use, the server should be **local**. The server will load on localhost, docker network OffRisk-net, port 8123.

This page receives off-target site location from the user in a tsv (tab delimiter file) format or written as text.

The tsv file should be in the following structure, each line is a site, and each value is separate with tab (chromosome, start position, end position, strand):

```
1 116905 116928 +
10 20041965 20041988 +
```

The text should be in the following structure, each line is a site, and each value is separate with a space:

1 116905 116928 +

10 20041965 20041988 +

Once you finish choosing the input click on **Run**, the results will be presented on the same page. For more information on the result please refer to the Result section.

On-target page

This page receives on-target site list from the user in json file format or filling the relevant fields.

The different options are:

pattern dna bulge: Relevant for Cas-OFFinder-bulge run. The desire DNA bulge size.

pattern rna bulge: Relevant for Cas-OFFinder-bulge run. The desire RNA bulge size.

<u>Sites:</u> A list of all desired sequences. Each site has the sequence and number of mismatches. The default number of mismatchs is 4. For FlashFry the sequence needs to be with PAM.

```
db list: A list defining the databases to analyze. The default is "all". Options databases are:
["gencode", "mirgene", "remapepd", "enhanceratlas", "pfam", "targetscan", "omim",
"humantf", "protein_atlas", "rbp", "cosmic"]
```

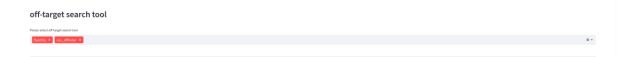
<u>search tools:</u> For on-target search, define which tool will search for off-target location – FlashFry or Cas-OFFinder-bulge. For more information on each please refer to the relevant guide. It is recommended to use FlashFry, since Cas-offinder can take longer and require more computational power

From UI:

Select the relevant server and database for this analysis.



Select search tool. The default is FlashFry.



Fill in the desired Pattern, DNA and RNA bulge, sequance and number of mismatch.



From file

The Json format needs to be in SiteList object structure. An example:

```
"flashfry"
]
}
Choose the file here. Must be a Json file.

Upload json file for on-target analysis

On-Target terreptual

Diag and drop file here

Diag and drop file here

On (Japan Json)

On (Japan Json)

On (Japan Json)

N

N

N
```

For more information on the result please refer to the Result section.

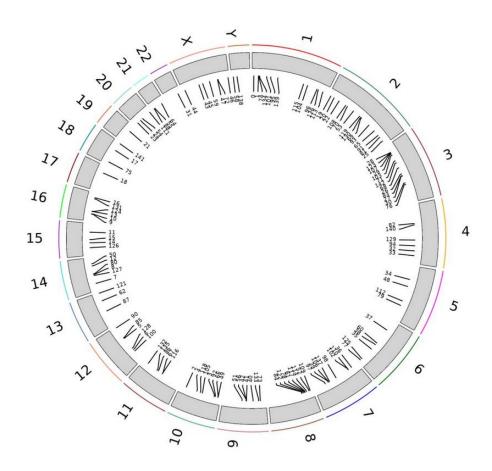
Result

The analysis results are shown as followed:

Circular view for off-target location in genome

This circular graph shows the result off-target locations on the different chromosomes

Circular view for off target location in genome



FlashFry score summary:

If FlashFry was chosen to run, the score result from it will be shown.

FlashFry Score Summary



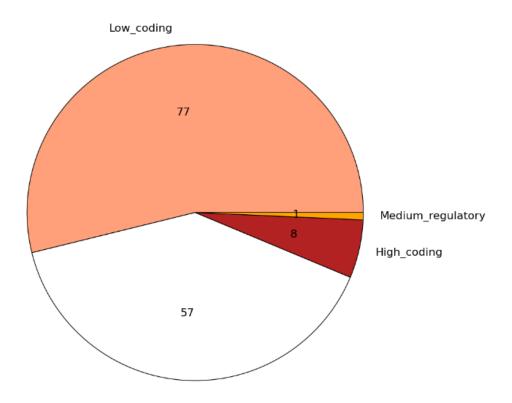
Off Target Summary

Off target database intersection result. Each off-target is colored according to the risk score that was calculated. If on-target mode was used with FlashFry, the information on the number of mismatch and occurrence will also be available.

Off-targets Summary



Score sum



GENCODE summary

This section contains the complete result from GENCODE intersection. First will be the complete result table, and then pie graph with the different distribution for gene type, segment that are coding and segment that are not.

Detailed biological information



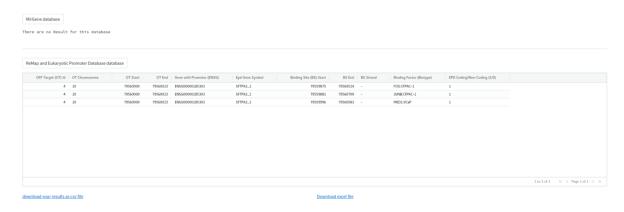
MirGene DB summary

This section contains the complete result from MirGene intersection.



ReMap & EPD summary

This section contains the complete result table from ReMap & EPD intersection.



Enhancer Atlas summary

This section contains the complete result table from Enhancer Atlas intersection



Pfam summary

This section contains the complete result table from Pfam intersection.



TargetScan summary

This section contains the complete result from table TargetScan intersection.



OMIM summary

This section contains the complete result table from OMIM intersection.



Human TF summary

This section contains the complete result table from Human TF intersection.



Protein Atlas summary

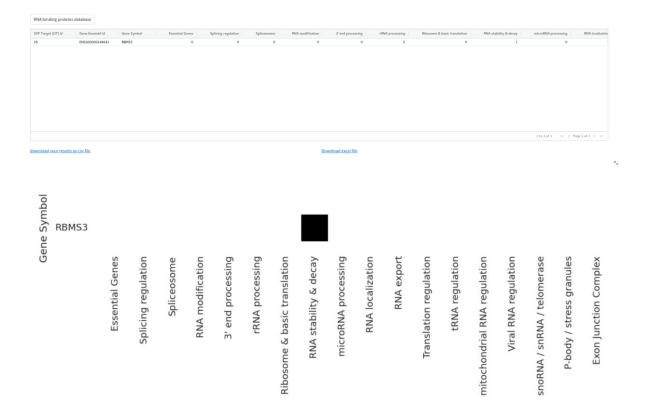
This section contains the complete result from Protein Atlas intersection. First will be presented the table, and then a heatmap colored according to the level of expression.





RBP Summary

This section contains the complete result table from RBP intersection. First will be the table, and then a heatmap showing the gene function.



COSMIC summary

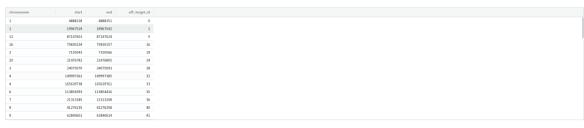
This section contains the complete result table from COSMIC intersection.



No result summary

All the off-targets that got no hit in the different database will be presented in the final table.

The following off-target ID did not match in any DB:



download your results as cov file Download excel file