

[Download and use a stand-alone version of RBPmap](#)

- [Requirements](#)
- [Download files](#)
- [Installation and configuration](#)
- [Usage](#)
- [Results](#)

Requirements

- Linux operating system
- Disk space: 20-30Gb (Depends on the number of genomes).
- Perl (including the following special modules: [DBI](#), [Statistics::Distributions](#) and [Math::NumberCruncher](#)).
- MySQL server (<http://dev.mysql.com/doc/refman/5.7/en/linux-installation.html>)

Download the following files:

1. RBPmap_1.2.tar.gz - RBPmap package (including RBPmap scripts and UCSC files and utilities)
2. RBPmap_1.2_MySQL_dump.sql - RBPmap MySQL dump file

Installation and configuration

Installation and configuration of RBPmap package:

1. Untar RBPmap package ('tar -zxvf RBPmap_1.2.tar.gz' will extract it to the current directory).
The package includes:
 - RBPmap scripts.
 - A sub-directory named 'RBP_PSSMs' which contains the PSSM motifs stored in RBPmap database (a file per each motif).
 - A sub-directory named 'UCSC' which contains several utilities needed for RBPmap run.
 - Under the 'UCSC' directory, a directory for each database assembly that is supported by RBPmap.
2. Add execution permissions ('chmod +x') to the following files:
 - [RBPmap.pl](#)
 - [UCSC/nibFrag](#)
 - [UCSC/faToNib](#)
 - [hgWiggle](#) (UCSC/hg19/hgWiggle, UCSC/hg18/hgWiggle, etc...)
3. Open the main script, [RBPmap.pl](#), and fill the variables at the top section with the correct paths of your environment:
 - [\\$scripts_dir](#) - An absolute path to the main directory of RBPmap (in which the scripts are located).
 - [\\$results_dir](#) - An absolute path to the main directory under which the results sub-directories will be created. Add writing permissions to this directory ('chmod +w').
4. Create a file named '.hg.conf' under your home directory (it is needed for the hgWiggle command, used when applying the conservation-based filtering). This file should contain the following lines :
[db.host=genome-mysql.cse.ucsc.edu](#)
[db.user=genomep](#)
[db.password=password central.db=hgcentral](#)

Create and configure RBPmap MySQL database:

1. Run the SQL script [RBPmap_1.2_MySQL_dump.sql](#) to create all RBPmap schemas and tables. (This can be done using the command-line or a designated tool, such as 'MySQL Workbench').
2. Create a file named [.my.cnf](#) under RBPmap main directory. This file should store the user and password parameters using to connect to MySQL. The [.my.cnf](#) file should be of the following

format:

[client]

user=<user>

password=<password>

Protect this file with the following permissions: `chmod 600 .my.cnf`.

- In `RBPmap.pl` script, fill the parameters required to connect to your MySQL server (stored in `%mysql` variable).

Install the desired genomes from UCSC genome browser:

The database assemblies, on which RBPmap search is performed, have to be installed locally. They should be located under the 'UCSC' directory (for example, under: `UCSC/hg19/`). Download and install the desired genomes as following:

- Go to UCSC FTP site: <ftp://hgdownload.cse.ucsc.edu/goldenPath/>
- Download the chromosomes in FASTA format (.fa.gz, only the regular ones).
For example: <ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/>
- Use the `faToNib` utility to translate the FASTA files (.fa) into NIB files (.nib), needed for RBPmap run.
For example: `faToNib chr1.fa chr1.nib`

Usage

```
perl RBPmap.pl -input <input file path> [options]
```

Input (mandatory):

- `-input <input file>`: A path (absolute or relative) to the input file, containing at least one sequence in genomic coordinates format: 'chromosome':start'-end':strand' ([view example](#)).

Optional flags:

- `-help`: Print RBPmap manual.

Genome Selection:

- `-genome <'human'/'mouse'/'drosophila'>`: The genome of the query sequences (default is 'human').
- `-db <'hg38'/'hg19'/'hg18'/'mm10'/'mm9'/'dm3'>`: The database assembly of the query sequences (should fit the genome).

Motif Selection:

- `-db_motifs <options>`: Select one or more motifs from RBPmap database (default is 'all_human' - all RBPmap motifs). Available options:
 - 'all': All RBPmap stored motifs.
 - 'all_human': All RBPmap human/mouse stored motifs.
 - 'all_drosophila': All RBPmap drosophila stored motifs.
 - `<protein1,protein2,...>`: Select all the motifs of the mentioned proteins names. The protein names should be separated by commas, without spaces, case insensitive. ([View RBPmap motifs list](#)).
 - For example: `-db_motifs HNRNPA1,HNRNPF,HNRNPH1`
 - 'none': None of RBPmap stored motifs (requires providing user-defined motifs).
- `-consensus <protein1:motif1,protein2:motif2,...>`: An option to provide one or more user-defined consensus motifs. Valid motifs are 4-10 characters long and contain IUPAC symbols only, no spaces between the motifs.
For example: `-consensus PTBP1:cucucu,MBNL1:ygcuky`
- `-pssm <probability matrix file path>`: An option to provide one or more user-defined PSSM motifs. The matrix file should be written in MEME format ([view example](#)).

Advanced Options:

- `-stringency <'high'/'default'>`: Control the stringency level of the results.
- `-conservation <'on'/'off'>`: Apply a conservation-based filter on intergenic regions (default is: 'off').

General Options:

- **-job_name <name>**: A name for the output directory of the job, without spaces (default is timestamp).
- **-delete_fasta <'on'/'off'>**: Delete the intermediate fasta files which are created during RBPmap run (default is 'on').

Results

The intermediate and output files of each RBPmap job will be created in a sub-directory named by the 'job_name' parameter, under the main results directory (defined in the variable \$results_dir inside RBPmap.pl).

- **All_Predictions.txt**: The main predictions summary file, containing the predictions for all the sequences submitted in the same job. It is located under the results directory created for the job.
- Output files for each sequence. These files are located under the sub-directories, created for each sequence (named 'sequence1', 'sequence2', etc...):
 1. **Predictions.txt**: A text file containing the predictions for the sequence.
 2. **RBPmap_custom_tracks.txt**: RBPmap binding sites predictions in BedGraph format. This file can be uploaded to UCSC Genome Browser to visualize the results as custom tracks.

For more information about RBPmap output files, see [RBPmap manual](#).

Website created by [Inbal Paz](#). Last update: 01/03/2021.