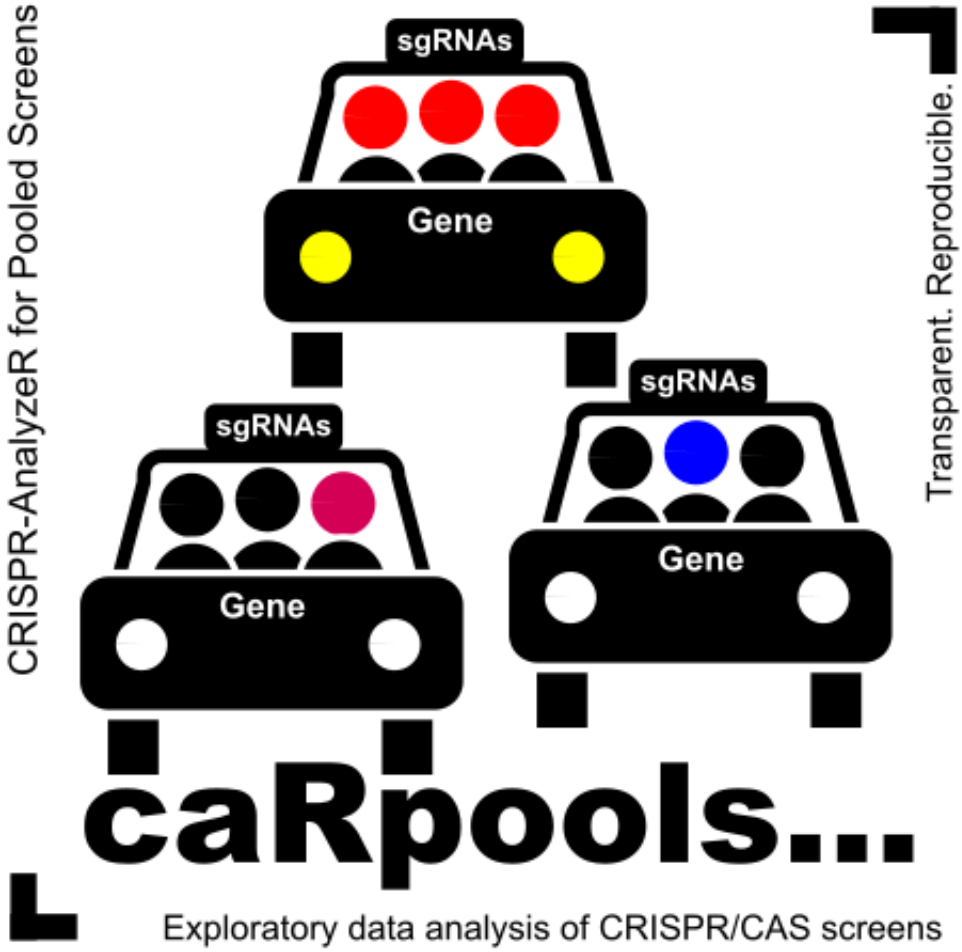
caRpools Shortcut User Guide

Jan Winter

Contents


CRISPR-AnalyzeR for Pooled Screens
sgRNAs
Gene
sgRNAs
Gene
sgRNAs
Gene
Transparent. Reproducible.
caRpools...
Exploratory data analysis of CRISPR/CAS screens

1 Files and Folder Structure to use CaRpools

Please note: the MAIN FOLDER must be the R working directory!
Data and Script paths can be adjusted in the MIACCS file.

The following files are neccessary to use CaRpools for report generation:

MIACCS.xls
Minimum Information About CRISPR/Cas Screens. This file needs to be filled out to provide all necessary informations about the screen.

R Markdown Template files
Either CaRpools-extended-PDF.rmd, CaRpools-PDF.rmd or CaRpools-extended-HTML.rmd or CaRpools-HTML.rmd. Is the template for report generation.

Data Files
Two replicates per Control and Treated. Can be FASTQ files OR already mapped, not normalized read count files.

CRISPR-mapping.pl
PERL script to map your extracted FASTQ files, if desired (as indicated in the MIACCS.xls)

CRISPR-extract.pl
PERL script to extract 20 nt target sequence from FAST files, if desired (as indicated in the MIACCS.xls)

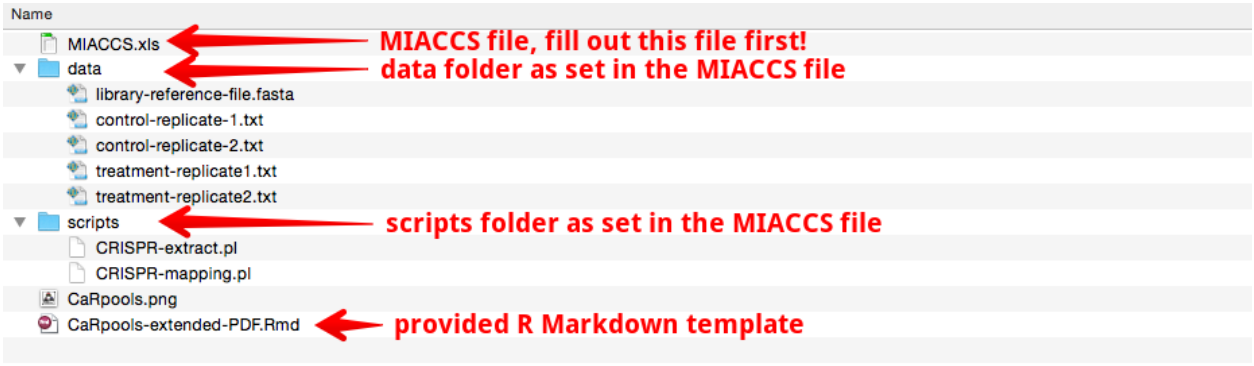
CaRpools.png
The logo file

The following files are necessary to use *single* CaRpools functions:

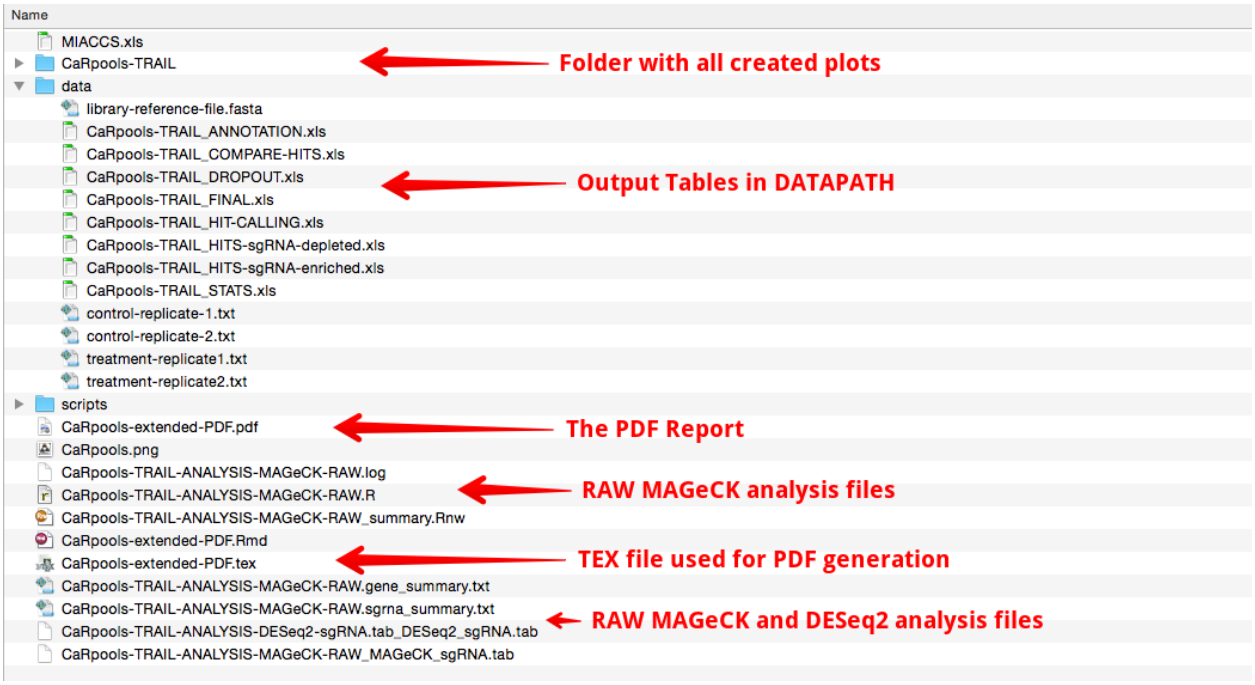
Data Files
Either raw read count files or FASTQ files (that need to be extracted and mapped using CaRpools)

Please note that CaRpools always starts with loading data files. For raw-readcount files, use `load.file`. For FASTQ files, please see the sections below.

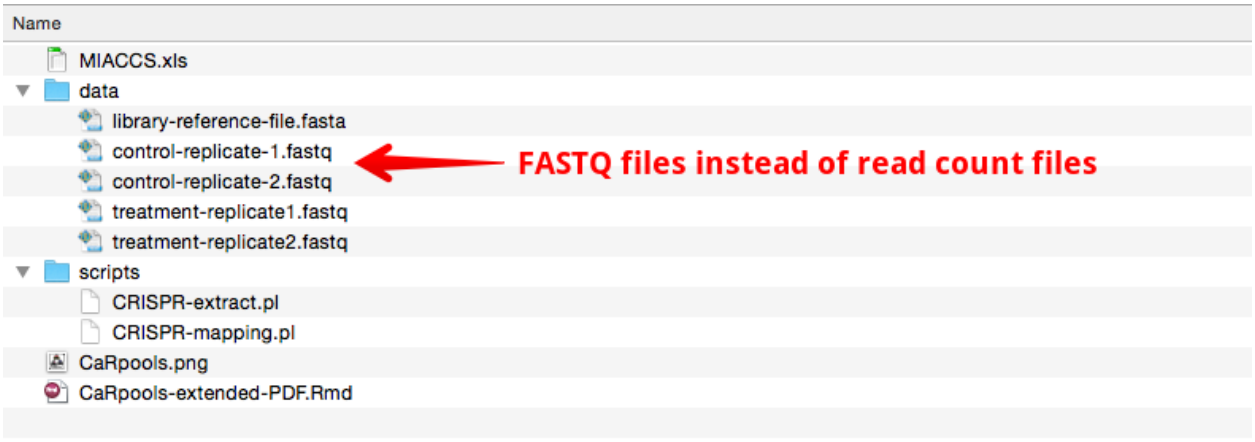
CaRpools folder structure for Report Generation using raw Read Count files:



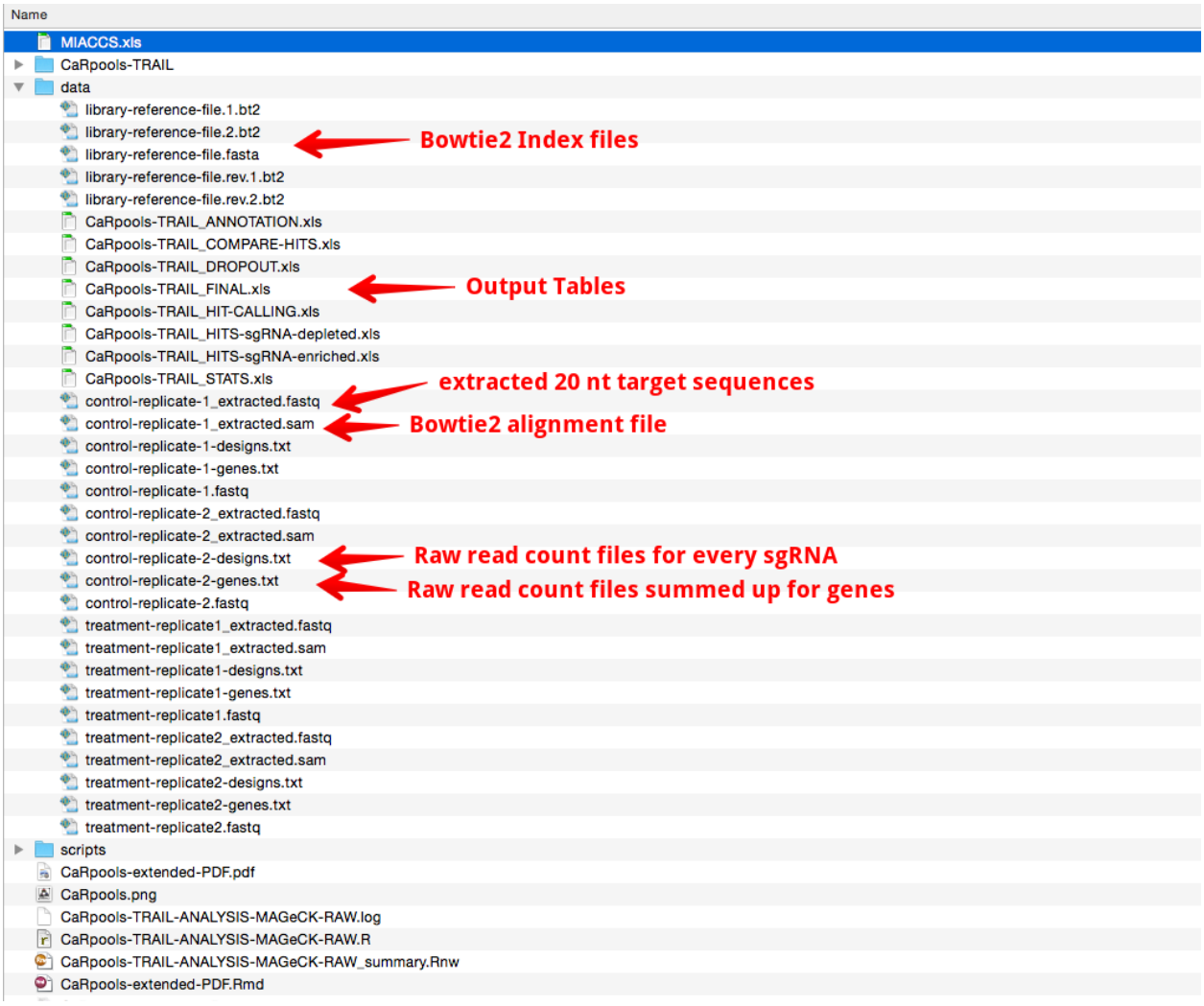
CaRpools folder structure for Report Generation using raw Read Count files AFTER REPORT GENERATION:



CaRpools folder structrue for Report Generation using FASTQ files:



CaRpools folder structure for Report Generation using FASTQ files AFTER REPORT GENERATION:


Name
MIACCS.xls
CaRpools-TRAIL
data
library-reference-file.1.bt2
library-reference-file.2.bt2
library-reference-file.fasta
Bowtie2 Index files
library-reference-file.rev.1.bt2
library-reference-file.rev.2.bt2
CaRpools-TRAIL_ANNOTATION.xls
CaRpools-TRAIL_COMPARE-HITS.xls
CaRpools-TRAIL_DROPOUT.xls
CaRpools-TRAIL_FINAL.xls
Output Tables
CaRpools-TRAIL_HIT-CALLING.xls
CaRpools-TRAIL_HITS-sgRNA-depleted.xls
CaRpools-TRAIL_HITS-sgRNA-enriched.xls
CaRpools-TRAIL_STATS.xls
extracted 20 nt target sequences
control-replicate-1_extracted.fastq
control-replicate-1_extracted.sam
Bowtie2 alignment file
control-replicate-1-designs.txt
control-replicate-1-genes.txt
control-replicate-1.fastq
control-replicate-2_extracted.fastq
control-replicate-2_extracted.sam
control-replicate-2-designs.txt
Raw read count files for every sgRNA
control-replicate-2-genes.txt
Raw read count files summed up for genes
control-replicate-2.fastq
treatment-replicate1_extracted.fastq
treatment-replicate1_extracted.sam
treatment-replicate1-designs.txt
treatment-replicate1-genes.txt
treatment-replicate1.fastq
treatment-replicate2_extracted.fastq
treatment-replicate2_extracted.sam
treatment-replicate2-designs.txt
treatment-replicate2-genes.txt
treatment-replicate2.fastq
scripts
CaRpools-extended-PDF.pdf
CaRpools.png
CaRpools-TRAIL-ANALYSIS-MAGeCK-RAW.log
CaRpools-TRAIL-ANALYSIS-MAGeCK-RAW.R
CaRpools-TRAIL-ANALYSIS-MAGeCK-RAW_summary.Rnw
CaRpools-extended-PDF.Rmd

1.1 Start CaRpools VirtualBox Image

Please see CaRpools-SHORTGUIDE-VirtualBox.pdf/html** or the shortguid here for how to use it.

1.2 Setup Files and R-Studio on your own computer

All packages and software tools need be installed correctly as shown before.

1. Copy all files in the designated folders as shown above.

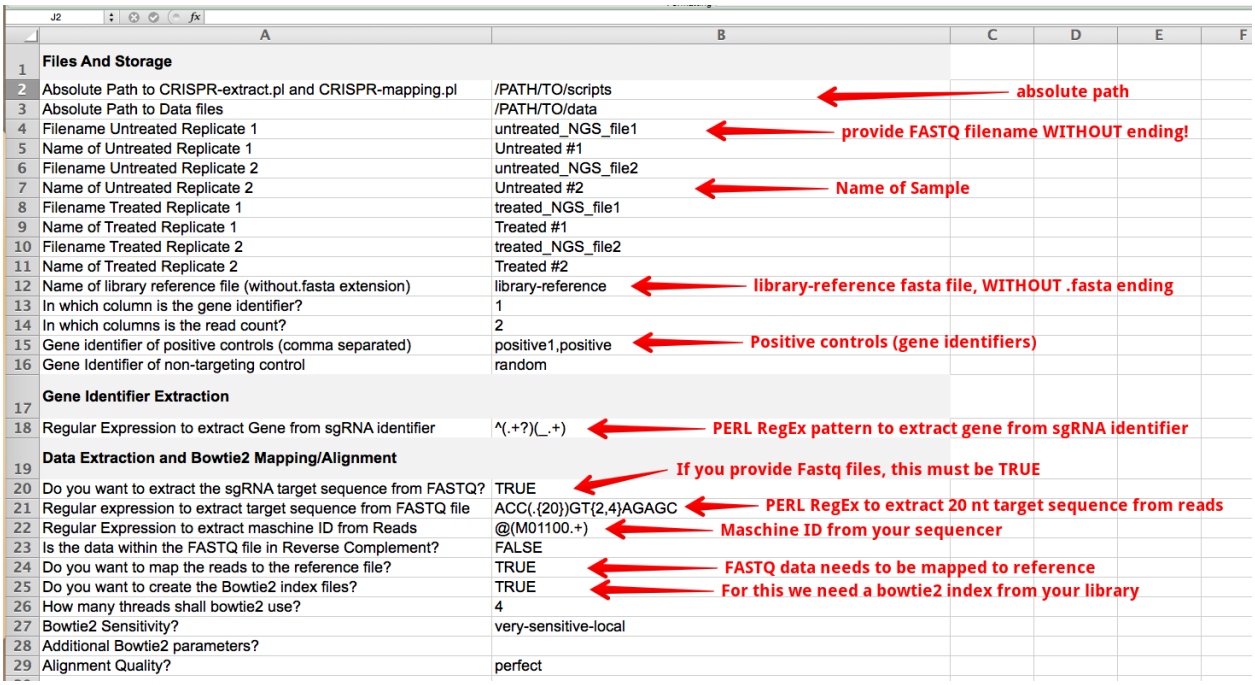
- **Please note: the MAIN FOLDER must be R working directory!**
- The MIACCS.xls as well the R markdown template and CaRpools.png must be in the same folder as the R working dir.

2. Adjust the path to the data and scripts folder if necessary in the MIACCS.xls . Use the absolute path. If the folder structure is as shown above, you do not need to make any adjustments.
3. Adjust and fill out the **MIACCS.xls** file.
4. You can use `CarPools(type="check")` to check for the correct folder structure and data file presence as it is indicated in the MIACCS.xls file.
5. You can check for your R working directory by `getwd()` and set it to any directory you want by `setwd("/PATH")`.

1.3 Check Setup

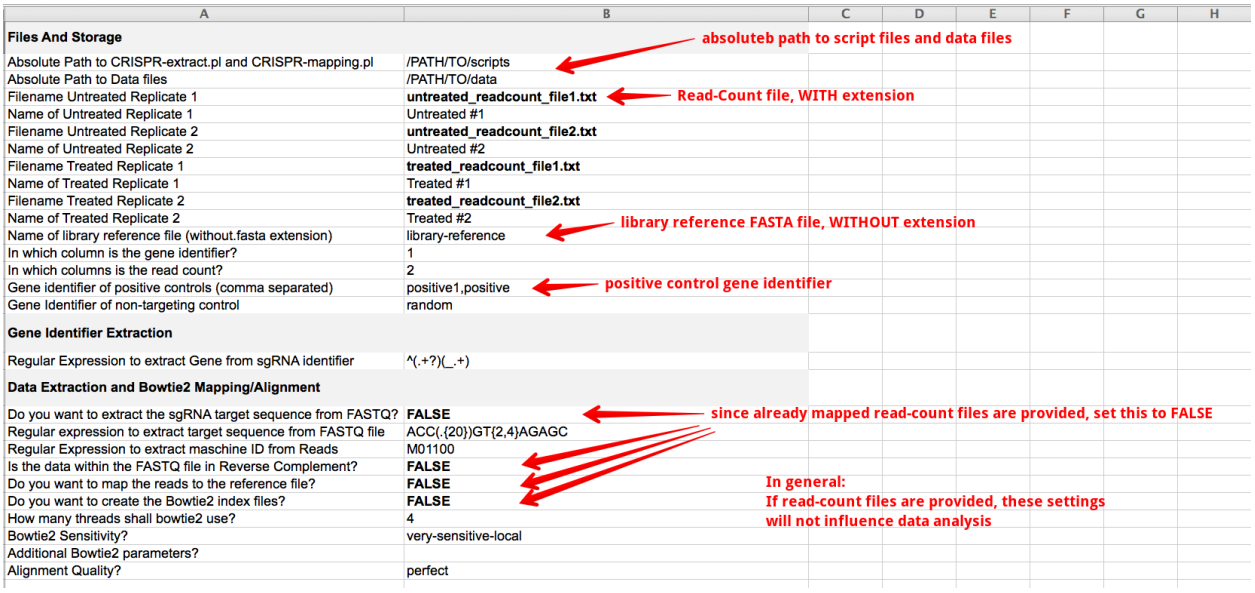
You can verify that the MIACCS.xls file as well as the used template file and all necessary scripts are found by calling `check.caRpools()`.
See below for more information about the arguments.
By default, it **requires a correct MIACCS file + the script files + all packages installed + MAGeCK + Bowtie2 + Pandoc.**

1.4 Example of a MIACCS File entry for FASTQ files



	A	B
1	**Files And Storage**	
2	Absolute Path to CRISPR-extract.pl and CRISPR-mapping.pl	/PATH/TO/scripts
3	Absolute Path to Data files	/PATH/TO/data
4	Filename Untreated Replicate 1	untreated_NGS_file1
5	Name of Untreated Replicate 1	Untreated #1
6	Filename Untreated Replicate 2	untreated_NGS_file2
7	Name of Untreated Replicate 2	Untreated #2
8	Filename Treated Replicate 1	treated_NGS_file1
9	Name of Treated Replicate 1	Treated #1
10	Filename Treated Replicate 2	treated_NGS_file2
11	Name of Treated Replicate 2	Treated #2
12	Name of library reference file (without.fasta extension)	library-reference
13	In which column is the gene identifier?	1
14	In which columns is the read count?	2
15	Gene identifier of positive controls (comma separated)	positive1,positive
16	Gene Identifier of non-targeting control	random
17	**Gene Identifier Extraction**	
18	Regular Expression to extract Gene from sgRNA identifier	^(.+?)(_.+)
19	**Data Extraction and Bowtie2 Mapping/Alignment**	
20	Do you want to extract the sgRNA target sequence from FASTQ?	TRUE
21	Regular expression to extract target sequence from FASTQ file	ACC(.{20})GT{2,4}AGAGC
22	Regular Expression to extract maschine ID from Reads	@(M01100.+)
23	Is the data within the FASTQ file in Reverse Complement?	FALSE
24	Do you want to map the reads to the reference file?	TRUE
25	Do you want to create the Bowtie2 index files?	TRUE
26	How many threads shall bowtie2 use?	4
27	Bowtie2 Sensitivity?	very-sensitive-local
28	Additional Bowtie2 parameters?	
29	Alignment Quality?	perfect

1.5 Example of a MIACCS File entry for Read-Count files



A	B
Files And Storage	
Absolute Path to CRISPR-extract.pl and CRISPR-mapping.pl	/PATH/TO/scripts
Absolute Path to Data files	/PATH/TO/data
Filename Untreated Replicate 1	**untreated_readcount_file1.txt**
Name of Untreated Replicate 1	Untreated #1
Filename Untreated Replicate 2	**untreated_readcount_file2.txt**
Name of Untreated Replicate 2	Untreated #2
Filename Treated Replicate 1	**treated_readcount_file1.txt**
Name of Treated Replicate 1	Treated #1
Filename Treated Replicate 2	**treated_readcount_file2.txt**
Name of Treated Replicate 2	Treated #2
Name of library reference file (without.fasta extension)	library-reference
In which column is the gene identifier?	1
In which columns is the read count?	2
Gene identifier of positive controls (comma separated)	positive1,positive
Gene Identifier of non-targeting control	random
Gene Identifier Extraction	
Regular Expression to extract Gene from sgRNA identifier	^(.+?)(_.+)
Data Extraction and Bowtie2 Mapping/Alignment	
Do you want to extract the sgRNA target sequence from FASTQ?	**FALSE**
Regular expression to extract target sequence from FASTQ file	ACC(.{20})GT{2,4}AGAGC
Regular Expression to extract maschine ID from Reads	M01100
Is the data within the FASTQ file in Reverse Complement?	**FALSE**
Do you want to map the reads to the reference file?	**FALSE**
Do you want to create the Bowtie2 index files?	**FALSE**
How many threads shall bowtie2 use?	4
Bowtie2 Sensitivity?	very-sensitive-local
Additional Bowtie2 parameters?	
Alignment Quality?	perfect

2 Start CaRpools Report Generation

You can start caRpools Report Generation after you did the following steps:

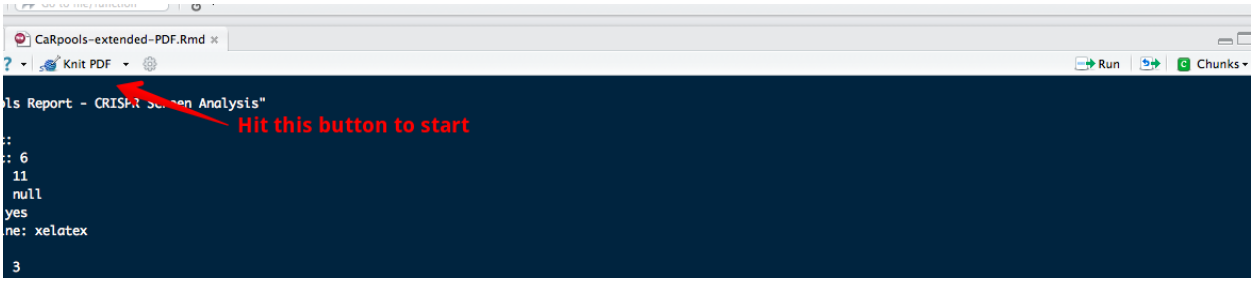
- Installed all required software and R packages (use *check.caRpools(files=FALSE)* to verify)

- Put every file in the correct folder (MIACCS, data files, script files, Rmd templates)
- Put everythin in the R working directory or set the working directory to the folder of your files
- Filled out the MIACCS file with all information, e.g. correct filenames, reference, data analysis options

You can check for all requirements by calling `check.caRpools`.

2.1 Start CaRpools using R-Studio

In the case you use R-Studio, you can start caRpools by just opening the corresponding Rmd template file. At the top, you will find the **Knit PDF** or **Knit HTML** button, so you just need to press that and caRpools will generate the report.



As an alternative, you can start caRpools via `use.caRpools` and provide additional parameters (see below).

2.2 Start CaRpools using R console

Moreover, caRpools report generation can also be initiated without R-studio installation, so that this can be done via R command line even on remote computers.
In this case, caRpools report generation can be started via `use.caRpools` with additional parameters, which are described below.

2.2.0.1 use.caRpools()

Usage:
use.caRpools(type=NULL, file="CaRpools-extended-PDF.Rmd", miaccs="MIACCS.xls", check=TRUE, work.dir=NULL)

type
Description If you provide a custom Rmd template that can generate both, PDF and HTML reports you can indicate which version you want to generate.
Default NULL
Values "PDF", "HTML"

file
Description The file name of your custom Rmd template file (with extension).
Default "CaRpools-extended-PDF.Rmd"
Values filename as character

miaccs
Description The filename of your MIACCS file.
Default "MIACCS.xls"
Values filename as character

check
Description Indicates whether caRpools will check for correct installation and file access.

Default TRUE
Values TRUE or FALSE (boolean)

work.dir
Description You can provide the absolute path to the working directory in which all files are placed (e.g. the MIACCS.xls and Rmd template).
Default NULL *Values* absolute path (character) or NULL if standard R working directory is used